

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:37:40 : Search time 276.658 Seconds
(without alignments)
9873.841 Million cell updates/sec

Title: US-09-831-426C-4

Perfect score: 1213
Sequence: 1 gtccgcgcgcgcgcgaag.....cagctactacccttgctaa 1213

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1213	100.0	1213	21	AAA15407	Fragment of DNA en
2	1211.4	99.9	1273	21	AAA15405	DNA encoding a hum
3	1206.2	99.4	1349	22	AAH34586	Human colon cancer
4	1171.6	96.6	1381	24	ABR83650	Human cDNA differe
5	1170	96.5	1399	17	AAT14038	Transcription fact
6	1167	96.2	1269	17	AAT14037	Transcription fact
7	495	40.8	690	24	ABO54923	Human ovarian anti
8	418.4	34.5	439	24	ABR44475	cDNA encoding colo
9	416.8	34.4	439	24	ABR45138	cDNA encoding colo

10	374	30.8	537	24	ABO59499	Human colon cancer
11	301.4	24.8	466	23	ABV49231	Human prostate exp
12	300.4	24.8	425	23	ABV19450	Human prostate exp
13	246	20.3	1401	22	AAH2129	Human eukaryotic 1
14	236	19.5	1889	23	AA573534	DNA encoding novel
15	234	19.3	234	24	ABK45059	cDNA encoding colo
16	193	15.9	291	16	AAT19568	Human gene signal
17	177	14.6	272	24	AA169173	Activated T-cell d
18	152.4	12.6	660	22	AAH23370	Nucleotide sequenc
19	152.4	12.6	947	22	AA511428	DNA encoding zinc
20	152.4	12.6	947	22	AAH89338	Plasmid pF11A21f
21	152.4	12.6	947	22	AAC91046	Vector TFI1A/Zif-
22	152.4	12.6	995	22	AA511427	DNA encoding zinc
23	152.4	12.6	995	22	AAH89337	Plasmid pF11A21f
24	152.4	12.6	995	22	AAC91045	Vector TFI1A/Zif-
25	101.6	8.4	183	24	ABN24558	Human ORF polyunc
26	97	8.0	919	22	AAH48456	Japanese medaka fe
27	91.2	7.5	3302	24	ABL69125	Kidney cancer rela
28	81	6.7	368	24	ABK46196	cDNA encoding colo
29	80.4	6.6	2400	22	ABA48579	Human breast cell
30	80.4	6.6	2400	22	ABA66494	Human foetal liver
31	80.4	6.6	2400	22	ABA33534	Probe #12020 for g
32	80.4	6.6	2400	22	AAK14917	Human brain expres
33	80.4	6.6	2400	22	AAK40649	Human bone marrow
34	80.4	6.6	2400	22	AA121411	Probe #11344 for g
35	80.4	6.6	2400	22	AA146696	Probe #15382 used
36	80.4	6.6	2400	22	AA107105	Probe #7096 used t
37	80.4	6.6	2400	24	AB514609	Human genome-deriv
38	78.8	6.5	2981	22	AAK53131	Human polyuncleoti
39	75.6	6.2	3135	22	AAK52147	Human polyuncleoti
40	74.6	6.2	1662	22	AAH34049	Human colon cancer
41	74.6	6.2	4017	23	ABV24337	Human prostate exp
42	74.2	6.1	1830	24	ABR69648	Mouse Sp1 family t
43	73.2	6.0	1197	24	ABR69649	Human Sp1 family t
44	67	5.5	1675	23	AA579321	DNA encoding novel
45	64.6	5.3	292	22	AA573932	DNA encoding zinc

ALIGNMENTS

RESULT 1
AAA15407
ID AAA15407 standard; DNA: 1213 BP.

AC AAA15407:

DT 04-SEP-2000 (first entry)

DE Fragment of DNA encoding a transcription factor designated htfl11A.

XX Human: transcription factor; htfl11A: DNA-binding protein;

KW transcription; ribosomal RNA 5S gene; transcriptional control;

KM cancer; ss.

XX Homo sapiens.

OS WO200028024-A1.

PN 18-MAY-2000.

PD 09-NOV-1999: 99WO-FR02738.

PF 10-NOV-1998: 98FR-0014146.

PR (HMRI) HOECHST MARION ROUSSEL.

PA Bordon-Pallier F, Roher C;

PI WPL: 2000-387419/33.

DR New nucleic acid encoding human transcription factor 11A, useful for

PT treatment and diagnosis of cancer and inherited disease

XX Claim 4: Page 42: 49pp: French.
PS
XX

The present sequence represents a fragment of a human transcription factor (designated htfIIIA) gene. The polypeptide is probably a DNA-binding protein probably involved in initiating transcription of the gene for ribosomal RNA 5S and maintaining the stability of CC the gene for other control genes. The htfIIIA polynucleotides and CC polypeptides are used to make therapeutic or diagnostic compositions for diseases associated with disorders of transcriptional control, CC particularly cancer or other inherited diseases. The htfIIIA CC polynucleotide can also be used to detect anomalies in gene CC transcription, particularly for diagnosis of inherited disease, also CC for studying diseases involving htfIIIA.
XX

SO Sequence 1213 BP: 349 A: 309 C: 302 G: 253 T: 0 other:

Query Match 100.0%; Score 1213: DB 21: Length 1213;
Best Local Similarity 100.0%; Pred. No. 2, 2e-314;
Matches 1213: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

```

QY 1 GTGCGGCGCGCGCGGAGGTTTCAGCAGGAGCGCTGGGCGCGCGGCTTCCCGGC 60
DB 1 GTGCGGCGCGCGCGGAGGTTTCAGCAGGAGCGCTGGGCGCGCGGCTTCCCGGC 60
QY 61 ACCTGTCTCGGACAGTGGACGCGCGCTGGGCGCTTGGAGGCGCGCGCTTGA 120
DB 61 ACCTGTCTCGGACAGTGGACGCGCGCTGGGCGCTTGGAGGCGCGCGCTTGA 120
QY 121 TCGCGCGCGCGTGGTGGCGGAGTGGTGTCTGCTTGCATGCGCGCGCTTCA 180
DB 121 TCGCGCGCGCGTGGTGGCGGAGTGGTGTCTGCTTGCATGCGCGCGCTTCA 180
QY 181 AGCGGCGGAGAGTCACTCCGAGCCCGCGCGCGCGCGCTTCCGCGAGTTCAT 240
DB 181 AGCGGCGGAGAGTCACTCCGAGCCCGCGCGCGCGCGCTTCCGCGAGTTCAT 240
QY 241 CTCTCTTCCCTGAGTGCAGCGCCCAATTACAGCAAAAGCTTGAAGCTTAC 300
DB 241 CTCTCTTCCCTGAGTGCAGCGCCCAATTACAGCAAAAGCTTGAAGCTTAC 300
QY 301 CAGGAGTACCATCTGAGCGCGCAGTTCGACTGACACAGAGAAAGCCGTTG 420
DB 301 CAGGAGTACCATCTGAGCGCGCAGTTCGACTGACACAGAGAAAGCCGTTG 420
QY 421 TGCAGCCACTGGCTGTGATCAAAATTCACACAAATTCAAATTCGAAGAA 480
DB 421 TGCAGCCACTGGCTGTGATCAAAATTCACACAAATTCAAATTCGAAGAA 480
QY 481 ACGCAACATGAAATATCAACAAATATATATGAGTTCGAGAGTGAAGAG 540
DB 481 ACGCAACATGAAATATCAACAAATATATATGAGTTCGAGAGTGAAGAG 540
QY 541 CTTTAAAGAACATGAGAGTGAAGAAATTCAGTGCAGATCCATGAACCT 600
DB 541 CTTTAAAGAACATGAGAGTGAAGAAATTCAGTGCAGATCCATGAACCT 600
QY 601 CAAGTGTACCCAGAGAGATGTGGAAACACTTTCATCACCAGCAAGCTGA 660
DB 601 CAAGTGTACCCAGAGAGATGTGGAAACACTTTCATCACCAGCAAGCTGA 660
QY 661 TGCAGAGCCCGACGAGGCTATGTATGTCAAAAAGATGTTCTTGTGG 720
DB 661 TGCAGAGCCCGACGAGGCTATGTATGTCAAAAAGATGTTCTTGTGG 720
QY 721 GACGGAATCTTGAACATGTGAGAGAAACCCATGAAGAGAAATGATGTA 780
DB 721 GACGGAATCTTGAACATGTGAGAGAAACCCATGAAGAGAAATGATGTA 780

```

```

QY 781 CCGGAAACATTTAAACGCAAGATTACCTTAAGACACATGAAAACTCCCGACA 840
DB 781 CCGGAAACATTTAAACGCAAGATTACCTTAAGACACATGAAAACTCCCGACA 840
QY 841 AAGGATGTATGTGCTGTTCACAGAGAGCTGTGGAAGACTATATCTGTTTAA 900
DB 841 AAGGATGTATGTGCTGTTCACAGAGAGCTGTGGAAGACTATATCTGTTTAA 900
QY 901 TCTCCAAAGCATATCTCTCTCTTCCATGAGAGAAAGCGCGCTTGTGTGAACAT 960
DB 901 TCTCCAAAGCATATCTCTCTCTTCCATGAGAGAAAGCGCGCTTGTGTGAACAT 960
QY 961 TGGCTGTGGCAAAACATTTGCAATGAAGAAAGTCTACTAGGCATGCTTGTATGA 1020
DB 961 TGGCTGTGGCAAAACATTTGCAATGAAGAAAGTCTACTAGGCATGCTTGTATGA 1020
QY 1021 TCTGTACAGAGAGAAATGAAGCTCAAAGTCAAAAAATCTGTGAAAAACGAGTTGGC 1080
DB 1021 TCTGTACAGAGAGAAATGAAGCTCAAAGTCAAAAAATCTGTGAAAAACGAGTTGGC 1080
QY 1081 CTCTCATCTCAGTGGATATATCCCTCCCAAGAAACAGGCGCAAGCTTATCTTGTG 1140
DB 1081 CTCTCATCTCAGTGGATATATCCCTCCCAAGAAACAGGCGCAAGCTTATCTTGTG 1140
QY 1141 TCAAAACGAGAGATCAACCACTGTGTGAAGACAAAGATGCTCTGACAGTTCAGTACT 1200
DB 1141 TCAAAACGAGAGATCAACCACTGTGTGAAGACAAAGATGCTCTGACAGTTCAGTACT 1200
QY 1201 TACCTTTGGCTAA 1213
DB 1201 TACCTTTGGCTAA 1213

RESULT 2
AA15405
ID AA15405 standard; DNA: 1273 BP.
XX
AC AA15405;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding a human transcription factor designated htfIIIA.
XX
KW Human; transcription factor; htfIIIA; DNA-binding protein;
transcription; ribosomal RNA 5S gene; transcriptional control;
cancer; SS.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 176..1273
FT FT /*tag= a
FT /product= "transcription factor"
XX
PD WO200028024-A1.
PD 18-MAY-2000.
PF 09-NOV-1999; 99MO-FR02738.
PR 10-NOV-1998; 98FR-0014146.
PA (HMRI ) HOECHST MARION ROUSSEL.
PI Bordon-Pallier F, Roher C;
XX
XX WPI; 2000-387419/33.
XX DR P-PDB: AAY93317.
XX PT New nucleic acid encoding human transcription factor IIA, useful for
XX PT treatment and diagnosis of cancer and inherited disease
XX PS Claim 3: Page 38-39: 49pp: French.

```

XX The present sequence encodes a human transcription factor designated CC hnf1a. The polypeptide is probably a DNA-binding protein probably involved in initiating transcription of the gene for ribosomal RNA 5S and maintaining the stability of transcription of other control genes. The hnf1a polynucleotides and polypeptides are used to make CC therapeutic or diagnostic compositions for diseases associated with CC disorders of transcriptional control, particularly cancer or other CC inherited diseases. The hnf1a polynucleotide can also be used to CC detect anomalies in gene transcription, particularly for diagnosis of CC inherited diseases, also for studying diseases involving hnf1a.

XX Sequence 1273 BP; 356 A; 329 C; 328 G; 260 T; 0 other:

Query Match 99.9%; Score 1211.4; DB 21; Length 1273;
Best Local Similarity 99.9%; Pred. No. 5.3e-314;
Matches 1212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 GTGCGGCGCGCGCGGAGGTTTCAGCAGAGGAGCGGTGGCGCGCGCGGTTCCCGCGC 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GTGCGGCGCGCGCGGAGGTTTCAGCAGAGGAGCGGTGGCGCGCGCGGTTCCCGCGC 120

QY 61 ACGTGTCTGCGCAGCTGGCGACCGCGCGCTGGGCTTGAGGCGCGCGCGCTTGA 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 ACGTGTCTGCGCAGCTGGCGACCGCGCGCTGGGCTTGAGGCGCGCGCGCTTGA 180

QY 121 TCGCGCGCGCGGTGGTGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TCGCGCGCGCGGTGGTGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240

QY 181 AGCGGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCTCCAGAGGTTCACTG 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AGCGGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCTCCAGAGGTTCACTG 300

QY 241 CTCCTTCCCTACTGTCAGCGCCCAATTACAGCAAAAGCTTGAAGCTTGACGCGCACCTG 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 CTCCTTCCCTACTGTCAGCGCCCAATTACAGCAAAAGCTTGAAGCTTGACGCGCACCTG 360

QY 301 CAACGACACGGGGGAGAGACATTGTTGTGACTATGAAAGGTGGCAGAGGCTTCAT 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CAACGACACGGGGGAGAGACATTGTTGTGACTATGAAAGGTGGCAGAGGCTTCAT 420

QY 361 CAGGACTACCATCTGAGCCGCCACATTTCTACTCACAGAGAGAAAAGCGTTTGTG 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 CAGGACTACCATCTGAGCCGCCACATTTCTACTCACAGAGAGAAAAGCGTTTGTG 480

QY 421 TGCAGCCACTGGCTGTGTATCAAAATTCACACAAATCAAACTTGAAGAAATTTGA 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TGCAGCCACTGGCTGTGTATCAAAATTCACACAAATCAAACTTGAAGAAATTTGA 540

QY 481 ACGCAACATGAAATGCAACAAATCAATATGACAGTTTGAAGACTTAAGAGAGC 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 ACGCAACATGAAATGCAACAAATCAATATGACAGTTTGAAGACTTAAGAGAGC 600

QY 541 CTTTAAAGAACATCAGCAGCGTGAATTCATCAGTGCAGATACCAATGAACTCTATT 600
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CTTTAAAGAACATCAGCAGCGTGAATTCATCAGTGCAGATACCAATGAACTCTATT 660

QY 601 CAAGTGTACCCAGAGAGATCTGGGAAACACTTTCATCACCCAGCAAGGTGAAGACA 660
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CAAGTGTACCCAGAGAGATCTGGGAAACACTTTCATCACCCAGCAAGGTGAAGACA 720

QY 661 TGCCAAAGCCACGAGGCTATGTATGTCAAAAGGATGTTCTTTGGGAAAAAATG 720
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 TGCCAAAGCCACGAGGCTATGTATGTCAAAAGGATGTTCTTTGGGAAAAAATG 780

QY 721 GACGGAATCTTGAAGATGTGAGAGAAACCATTAAGAGAGAAATAGTATGAGATG 780
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 GACGGAATCTTGAAGATGTGAGAGAAACCATTAAGAGAGAAATAGTATGAGATG 840

QY 781 CCGGAAAGCATTTAAAGCAAGATTACCTTAAGCAACACATGAAACTCATGCCGAGA 840
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 CCGGAAAGCATTTAAAGCAAGATTACCTTAAGCAACACATGAAACTCATGCCGAGA 900

```

```

QY 841 AAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGGAAGAACTTACTACTGTTTAA 900
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 AAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGGAAGAACTTACTACTGTTTAA 960

QY 901 TCTCCAAAGCCATATCCCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 960
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 TCTCCAAAGCCATATCCCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020

QY 961 TGGCTGTGCGAAACATTTGCAATGAAACAAAGTCTCACTAGCATGCTGTTGACATGA 1020
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 TGGCTGTGCGAAACATTTGCAATGAAACAAAGTCTCACTAGCATGCTGTTGACATGA 1080

QY 1021 TCCGACAGAGAGAAATGAGAGCTCAAGTCAAAAAATCTGTGAAAACGAGTTGGC 1080
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 TCTGTACAAAGAGAGAAAGAGAGCTCAAAATCTGTGAAAACGAGTTGGC 1140

QY 1081 CTCATCATCTCAGTGATATATCCCTCCCAAAAGGAAACAGGCGTATGCTTGTG 1140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 CTCATCATCTCAGTGATATATCCCTCCCAAAAGGAAACAGGCGTATGCTTGTG 1200

QY 1141 TCAAAACGAGAGATCACCCCACTGTGTGGAAGAGAGATGCTCGACAGTTGACATCT 1200
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 TCAAAACGAGAGATCACCCCACTGTGTGGAAGAGAGATGCTCGACAGTTGACATCT 1260

QY 1201 TACCTTGGCTTAA 1213
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 TACCTTGGCTTAA 1273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
AAH34586
ID AAH34586 standard; cDNA: 1349 BP.
XX
AC AAH34586;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1668.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
   colorectal carcinoma; chromosome 13; ss.
XX
OS Homo sapiens.
XX
PN W0200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000MO-US26524.
XX
PR 29-SEP-1999; 990S-0157137.
XX
PR 03-NOV-1999; 990S-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
XX
DR P-PSDB; AAG75181.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
   useful for preventing, diagnosing and/or treating colorectal cancers -
   Claim 1; Page 3273-3274; 9803p; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
   cancer-associated nucleic acid molecules (N) and proteins (P), where
   the proteins are collectively known as colon cancer antigens. The colon
   cancer antigens have cytostatic activity and can be used in gene
   therapy and vaccine production. N and P may be used in the prevention,
   diagnosis and treatment of diseases associated with inappropriate P
   expression. For example, N and P may be used to treat disorders

```

CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of p by expressing
 CC inactive proteins or to supplement the patients own production of p.
 CC Additionally, N may be used to produce the colon cancer-associated ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and p can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 1349 BP; 400 A; 337 C; 325 G; 285 T; 2 other:

Query Match 99.4%; Score 1206.2; DB 22; Length 1349;
 Best Local Similarity 99.8%; Pred. No. 1.3e-312;
 Matches 1208; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GCCGCGCCCGCGGAGGTTTCAGCAGGAGCCGCGCGCGCGCGGTTCCCGCGCAC 62
 DB 22 GCGCGCGCGCGCGGAGGTTTCAGCAGGAGCCGCGCGCGCGCGGTTCCCGCGCAC 81
 QY 63 GTGTCTCGCAGCTGGCAGCGCGCGCTGGGCTTGGAGCGCGCGCGCGCTGGATC 122
 DB 82 GTGTCTCGCAGCTGGCAGCGCGCGCTGGGCTTGGAGCGCGCGCGCGCTGGATC 141
 QY 123 CGCGCGCGCGCGCGGAGGTTTCAGCAGGAGCCGCGCGCGCGGTTCCCGCGCAC 182
 DB 142 CGCGCGCGCGCGCGGAGGTTTCAGCAGGAGCCGCGCGCGCGGTTCCCGCGCAC 201
 QY 183 CGCGCGAGAGCTAGCTCCGACCCCGCGCGCGCGCGGTTCCCGAGAGTTTCATGCT 242
 DB 202 CGCGCGAGAGCTAGCTCCGACCCCGCGCGCGCGCGGTTCCCGAGAGTTTCATGCT 261
 QY 243 CCTTCCCTGACTGAGCGCCCAATTACAGCAAAAGCTTGAAGCGCGCACCTGTGCA 302
 DB 262 CCTTCCCTGACTGAGCGCCCAATTACAGCAAAAGCTTGAAGCGCGCACCTGTGCA 321
 QY 303 AGCAGCGGGGAGAGACATTTGTTGTGACTGATGAAAGGTGGGAGAGGCTTCATCA 362
 DB 322 AGCAGCGGGGAGAGACATTTGTTGTGACTGATGAAAGGTGGGAGAGGCTTCATCA 381
 QY 363 GGGAGCTACATCTGAGCGCGCACATTTCTGACTCAGCAGAGAGAAACCGTTTGTG 422
 DB 382 GGGAGCTACATCTGAGCGCGCACATTTCTGACTCAGCAGAGAGAAACCGTTTGTG 441
 QY 423 CAGGCACTGGCTGTGATCAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGAAC 482
 DB 442 CAGGCAATGGCTGTGATCAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGAAC 501
 QY 483 GCAAAACATGAAATCAAAACAAATATATATGACATTTTGAAGACTGTGAAGAAACCT 542
 DB 502 GCAAAACATGAAATCAAAACAAATATATATGACATTTTGAAGACTGTGAAGAAACCT 561
 QY 543 TTAAAGAAACATCAGCAGCTGAAATTCATCAGTCCAGCATATCAATCAACCTTATTTCA 602
 DB 562 TTAAAGAAACATCAGCAGCTGAAATTCATCAGTCCAGCATATCAATCAACCTTATTTCA 621
 QY 603 AGTGTACCCAGAGAGATGTGGGAAACACTTTGCATCACCAGCAACCTGAAAGACATG 662
 DB 622 AGTGTACCCAGAGAGATGTGGGAAACACTTTGCATCACCAGCAACCTGAAAGACATG 681
 QY 663 CCAAGGGCCAGAGGGCATGTATGTCAAAAAGATGTCTCTTGTGGCAAAAACATGGA 722
 DB 682 CCAAGGGCCAGAGGGCATGTATGTCAAAAAGATGTCTCTTGTGGCAAAAACATGGA 741
 QY 723 CGGAACCTTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGCC 782
 DB 742 CGGAACCTTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGCC 801
 QY 783 GGAACATTTTAAAGCAAAAGATTACCTTAAAGCAACATGAAAACTCATGCCCCAGAAA 842

DB 802 GCAAAACATTTTAAAGCAAAAGATTACCTTAAGCNAACATGAAAACTCATGCCCCAGAAA 861
 QY 843 GGAATGATGTCCGTGTCCAAAGAAAGCTGTGGAACAACTACTAGTGTATATC 902
 DB 862 GGGATGATGTCCGTGTCCAAAGAAAGCTGTGGAAGAACTATACAACTGTGTTAATC 921
 QY 903 TCCAAAGCCATATCTCTTCCATGAGGAAAGCCGCGCTTGTGTGAAACATGCTG 962
 DB 922 TCCAAAGCCATATCTCTCTCCATGAGGAAAGCCGCGCTTGTGTGAAACATGCTG 981
 QY 963 GCTGTGGCAAAACATTTGGCAATGAACAAAGTCCATAGCATGTGCTCATATGATC 1022
 DB 982 GCTGTGGCAAAACATTTGGCAATGAACAAAGTCCATAGCATGTGCTCATATGATC 1041
 QY 1023 CTGCAAGAAAGAAATGAACTCAAAAGTCAAAAATTCGTGGAAGAAAGAGTTGGCCT 1082
 DB 1042 CTGCAAGAAAGAAATGAACTCAAAAGTCAAAAATTCGTGGAAGAAAGAGTTGGCCT 1101
 QY 1083 CTCATCTCAGTGTATATCCCTCCCAAAAGAAAGGCAAGGCTTATCTTGTGTC 1142
 DB 1102 CTCATCTCAGTGTATATCCCTCCCAAAAGAAAGGCAAGGCTTATCTTGTGTC 1161
 QY 1143 AAAAGGAGAGTCAACCACTGTGTGGAAGCAAGATGCTGTGACAGTTGCACTTA 1202
 DB 1162 AAAAGGAGAGTCAACCACTGTGTGGAAGCAAGATGCTGTGACAGTTGCACTTA 1221
 QY 1203 CCTTGGCTAA 1213
 DB 1222 CCTTGGCTAA 1232

RESULT 4
 ABR83650
 ID ABR83650 standard: cDNA: 1381 BP.

XX AC ABR83650;
 XX AC
 XX DT 14-AUG-2002 (first entry)
 XX XX

DE Human cDNA differentially expressed in granulocytic cells #221.

KW Human; ss: granulocytic cell; DNA chip: bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

PN NO200228999-A2.

PD 11-APR-2002.

PF 03-OCT-2001; 2001MO-US30821.

PR 03-OCT-2000; 2000US-237189P.

PA (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

DR WPI: 2002-435328/46.

PT Detecting granulocyte activation by detecting differential expression
 of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 drug toxicity

PS Claim 1; SEQ ID NO 221; 114pp; English.

CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) especially
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GCA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal infection, also bacterial infection, viral infection,
CC peristaltic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1381 BP; 386 A; 350 C; 354 G; 291 T; 0 other;

Query Match 96.6%; Score 1171.6; DB 24; Length 1381;
Best Local Similarity 99.4%; Pred. No. 2,4e-303;
Matches 1207; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 1 GTGCGGCGCGCGCGGAGGTTTCAGACAGGAGCGGTGGGCGCGCGGTTCCCGGCG 60
DB 80 GTGCGGCGCGCGCGGAGGTTTCAGACAGGAGCGGTGGGCGCGCGGCGCGCGCG 138
QY 61 ACGTGTCTGGCAGGTGGCAGCGCGCGCTGGGCTTGAGAGCGCGCGCGCTTGA 120
DB 139 ACGTGTCTGGCAGGTGGCAGCGCGCGCTGGGCTTGAGAGCGCGCGCGCTTGA 198
QY 121 TCCGCGCGCGCGCGTGGTCCGCGAGTGGTGTCTCTTCACATCGCGCGCGCTTATGGC 180
DB 199 TCCGCGCGCGCGCGTGGTCCGCGAGTGGTGTCTCTTCACATCGCGCGCGCTTATGGC 258
QY 181 AGCGGCGAGAGCTAGCTCGACCGCGCGCGCGCGCGCTTCCAGAGGTTTCATCTG 240
DB 259 AGCGGCGAGAGCTAGCTCGACCGCGCGCGCGCGCGCTTCCAGAGGTTTCATCTG 318
QY 241 CTCCTTCCCTAGCTAGCAGCGCGCAATTACAGCAAAAGCCTGGAAGCTTGAAGCGCACTGTG 300
DB 319 CTCCTTCCCTAGCTAGCAGCGCGCAATTACAGCAAAAGCCTGGAAGCGCACTGTG 378
QY 301 CAAGCAGACGCGGCGAGAGACATTTGTTGTGACTATGAAGGCTGGCAAGGCGCTTAT 360
DB 379 CAAGCAGACGCGGCGAGAGACATTTGTTGTGACTATGAAGGCTGGCAAGGCGCTTAT 438
QY 361 CAGGAGCTACATCTGAGCGCGCACATTTCTGACACAGAGAGAAAGCCGTTTGTG 420
DB 439 CAGGAGCTACATCTGAGCGCGCACATTTCTGACACAGAGAGAAAGCCGTTTGTG 498
QY 421 TGACGCCCATGCTGTGATCAAAAATTCACAAAATCAAACTGGAAGAAACATTTTGA 480
XX

DB 499 TGACGCCCATGCTGTGATCAAAAATTCACAAAATCAAACTGGAAGAAACATTTTGA 558
QY 481 ACGGAACATATAAATATCAACAAAAACAAATATATATGAGTTTGAAGCTTAAGAAC 540
DB 559 ACGGAACATATAAATATCAACAAAAACAAATATATATGAGTTTGAAGCTTAAGAAC 618
QY 541 CTTTAAGAAATATGAGCGGTGAATAATCATGAGCGAGATACCAATGAACGCTATT 600
DB 619 CTTTAAGAAATATGAGCGGTGAATAATCATGAGCGAGATACCAATGAACGCTATT 678
QY 601 CAAGTGTACCCAGAGAGATGAGGAAACACTTTCATACCCAGCAAGCTGAAGACACA 660
DB 679 CAAGTGTACCCAGAGAGATGAGGAAACACTTTCATACCCAGCAAGCTGAAGACACA 738
QY 661 TGCCAAAGGCCACGAGGCTTATGATGCAAAAAGATGTTCTTTGTGGCAAAACATG 720
DB 739 TGCCAAAGGCCACGAGGCTTATGATGCAAAAAGATGTTCTTTGTGGCAAAACATG 798
QY 721 GACGGAACCTTGAACACATGTGAGGAAACCATTAAAGAGAAATACATGTGAAGTATG 780
DB 799 GACGGAACCTTGAACACATGTGAGGAAACCATTAAAGAGAAATACATGTGAAGTATG 858
QY 781 CCGGAACCATTTAAACGCAAAAGATTACCTTAAGCAACACATGAACATCGCCCGAGA 840
DB 859 CCGGAACCATTTAAACGCAAAAGATTACCTTAAGCAACACATGAACATCGCCCGAGA 918
QY 841 AAGGATGTATGTGCTGTCTCAAGAGAGGCTGTGGAAGAACTATACCTACTGTGTTAA 900
DB 919 AAGGATGTATGTGCTGTCTCAAGAGAGGCTGTGGAAGAACTATACCTACTGTGTTAA 978
QY 901 TCTCCAAAGCATATCCCTCTCCATGAGGAAAGCGCCCTTTGTGTGAACATGCG 960
DB 979 TCTCCAAAGCATATCCCTCTCTCCATGAGGAAAGCGCCCTTTGTGTGAACATGCG 1038
QY 961 TGCGTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGCATGCTGTTTACATGA 1020
DB 1039 TGCGTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGCATGCTGTTTACATGA 1098
QY 1021 TCTGTGACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGTGTGAAGAAC- GAGTTTGG 1079
DB 1099 TCTGTGACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGTGTGAAGAACCGGAGTTTGG 1158
QY 1080 CCTCTCATCTAGTGAATATATTCCTCTCCAAAGAAACAAAGGCGCAAGGCTTATCTTGT 1139
DB 1159 CCTCTCATCTAGTGAATATAT- CCTCTCCAAAGAAACAAAGGCGCAAGGCTTATCTTGT 1217
QY 1140 GTCAAAACGAGAGTCAACCAACTGTGTGAAGACAAGATGCTCTCGACAGTTGCAGTAC 1199
DB 1218 GTCAAAACGAGAGTCAACCAACTGTGTGAAGACAAGATGCTCTCGACAGTTGCAGTAC 1277
QY 1200 TTACCTTGGCTAA 1213
DB 1278 TTACCTTGGCTAA 1291
RESULT 5
AAT14038
ID AAT14038 standard; cDNA: 1399 BP.
XX
AC AAT14038;
XX
DT 07-JUL-1996 (first entry)
XX
DE Transcription factor-IIIA gene.
XX
KW Human; transcription factor-IIIA; hNFIIIA; DNA binding protein;
KW ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;
KW primer; PCR; polymerase chain reaction; foetal brain;
KW anchor primer; diagnostic; probe; transcription control;
XX
OS Homo sapiens.
XX

Key Location/Qualifiers
FH CDS 20..1291
FT /*tag= a
FT /product= Transcription factor-IIIA
FT /note= "Claim 2"
FT misc-feature 317..1096
FT /*tag= b
FT /product= Zinc finger domains
FT polyA_signal 1363..1368
FT /*tag= c
XX EP704526-A1.
XX 03-APR-1996.
XX 05-SEP-1995: 95SEP-0113908.
XX 05-SEP-1994: 94UP-0211022.
XX 05-SEP-1994: 94UP-0211022.
XX (SAKA) OTSUKA PHARM CO LTD.
XX Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S:
XX WPI, 1996-173033/18.
XX P-PSDB: AAR91305.
XX Human Transcription Factor III A gene - useful in regulation of
PT transcription and for diagnosis and treatment of e.g. cancer related
PT diseases
XX
XX Example 1: Page 11-13; 17pp; English.
XX
CC The sequence encodes human transcription factor-IIIA (HNF1IA), a
CC DNA binding protein with 9 zinc finger domains, which is necessary
CC for the initiation of 5S RNA gene transcription, binding to an
CC internal control region of the 5S gene. The coding region
CC (claimed) is given in AAT14037. A fragment lacking a 5'-portion of
CC the gene has been isolated from a human foetal brain cDNA library
CC (OTK7-1), and the 5'-portion of the gene has been isolated by
CC 5'-rapid amplification of cDNA ends using primers H11-R (AAT14039),
CC H11-E (AAT14040), H11-H (AAT14041) and Ap-2 (AAT14042), complementary to
CC anchor primer AAT14043. Reverse transcription using H-11R is
CC followed by anchor primer ligation and PCR using Ap-2 and H11-E, to
CC give a full-length cDNA, OTK7. The gene and its encoded protein
CC may be used in diagnosis, identification or therapy of hereditary
CC diseases such as cancer, or other diseases resulting from abnormal
CC transcriptional control, and to analyse the mechanisms involved in
CC their activity.
XX
XX Sequence 1399 BP; 405 A; 349 C; 354 G; 291 T; 0 other;
SQ
Query Match 96.5%: Score 1170; DB 17: Length 1399;
Best Local Similarity 99.3%: Pred. No. 6.5e-303;
Matches 1206: Conservative 0; Mismatches 5; Indels 3; Gaps 3;
QY 1 GTCCGCGCCCGCCGGAAGTTTCAGACAGAGCCGTGGCGCGCGCCGTTCCCGGC 60
DB 80 GTCCGCGCGTCCGCGGAAGTTTCAGACAGAGCCGTGGCGCGCGCGG -CGCTTCCCGGC 138
QY 61 AGCTGTCTCGGACGTGGACAGCCGCGCTGGCGCTTTCAGAGGCCCGCGCGCTTGA 120
DB 139 ACGTGTCTCGGACGTGGACAGCCGCGCTGGCGCTTTCAGAGGCCCGCGCGCTTGA 198
QY 121 TCCGCGCGCGCGTGGTCCGCGAGTGGGTGTGCTCTTTCAGACATGCCGACAGGTTTCAT 180
DB 199 TCCGCGCGCGCGTGGTCCGCGAGTGGGTGTGCTCTTTCAGACATGCCGACAGGTTTCAT 258
QY 181 AGCGGCGAGAGCTACGTCGACGCCCGCGCGCGCGCTTTCAGAGAGTTTCATCTG 240
DB 259 AGCGGCGAGAGCTACGTCGACGCCCGCGCGCGCGCTTTCAGAGAGTTTCATCTG 318
QY 241 CTCCTTCCCTGACTGACAGCGCAATTACAGCAAGCTTGAAGCTTACAGCGCACTGTG 300
|||||

DB 319 CTCCTTCCCTGACTGACAGCGCAATTACAGCAAGCCCTGGAAGCTTACGCCGCACTGTG 378
QY 301 CAAGCACACGGGGAGAGACCATTTGTTGACTATGAAGGGTGTGGCAAGGCTTCAT 360
DB 379 CAAGCACACGGGGAGAGACCATTTGTTGACTATGAAGGGTGTGGCAAGGCTTCAT 438
QY 361 CAGGACTACCATTCGTGAGCCGCCACATTCCTGACTCAGACAGAGAGAAAGCGTTTGT 420
DB 439 CAGGACTACCATTCGTGAGCCGCCACATTCCTGACTCAGACAGAGAGAAAGCGTTTGT 498
QY 421 TGCAGCCACTGGTGTATCAAAAATTCACACAAATCAAACTTGAAGAAACATTTTGA 480
DB 499 TGCAGCCAAATGGTGTATCAAAAATTCACACAAATCAAACTTGAAGAAACATTTTGA 558
QY 481 AGCGAAACATGAAATACAAAATCAAAATATATGAGTTTGAACAGTGAAGAGAC 540
DB 559 AGCGAAACATGAAATACAAAATCAAAATATATGAGTTTGAACAGTGAAGAGAC 618
QY 541 CTTTAAAGAACATCAGCAGCTGAAATTCATCACTGCGCAGCATACCAATGAACCTTATT 600
DB 619 CTTTAAAGAACATCAGCAGATGAAATTCATCACTGCGCAGATACCAATGAACCTTATT 678
QY 601 CAAGTATACCCAGAGAGATGTGGAAACACTTTGCATCACCCAGCAAGCTGAACAGCA 660
DB 679 CAAGTATACCCAGAGAGATGTGGAAACACTTTGCATCACCCAGCAAGCTGAACAGCA 738
QY 661 TGCCAAAGCCAGAGGCTATGATGCAAAAAGATGTTCTTGTGGCAAAAACATG 720
DB 739 TGCCAAAGCCAGAGGCTATGATGCAAAAAGATGTTCTTGTGGCAAAAACATG 798
QY 721 GACGGAACCTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATTAATGTGAAGTATG 780
DB 799 GACGGAACCTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATTAATGTGAAGTATG 858
QY 781 CCGGAAACATTTTAACGCAAAATTTACCTTAAGCAACATGAATCTATGCCCGACA 840
DB 859 CCGGAAACATTTTAACGCAAAATTTACCTTAAGCAACATGAATCTATGCCCGACA 918
QY 841 AAGGATGATATGCGGTGTCACAGAAAGGCTGGAAGAAACCTATCTGTGTGTTAA 900
DB 919 AAGGATGATATGCGGTGTCACAGAAAGGCTGGAAGAAACCTATCTGTGTGTTAA 978
QY 901 TCTCCAAAGCCATTCCTTCCTTCATGAGAAAGCCGCTTTGTGTGTAACATGC 960
DB 979 TCTCCAAAGCCATTCCTTCCTTCATGAGAAAGCCGCTTTGTGTGTAACATGC 1038
QY 961 TGGCTGTGGCAAAACATTTGCAATGAAGAAAGTCTCACTAGGATCTGTGTACATGA 1020
DB 1039 TGGCTGTGGCAAAACATTTGCAATGAAGAAAGTCTCACTAGGATCTGTGTACATGA 1098
QY 1021 TCTGACAAAGAAATGAAGTCAAAAGTCAAAATCTCGTGAAGAAC -GAGGTTGG 1079
DB 1099 TCTGACAAAGAAATGAAGTCAAAAGTCAAAATCTCGTGAAGAAAGGAGTTGG 1158
QY 1080 CCTCTCATCTCAGTATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTGT 1139
DB 1159 CCTCTCATCTCAGTATATAT -CCTCCAAAGAAACAAAGGCAAGGCTTATCTTTGT 1217
QY 1140 GTCAAAACGAGAGTCAACCAACTGTGTGAAGCAAGAGTCTGTGACAGTTCAGTAC 1199
DB 1218 GTCAAAACGAGAGTCAACCAACTGTGTGAAGCAAGAGTCTGTGACAGTTCAGTAC 1277
QY 1200 TTACCTTGGCTAA 1213
DB 1278 TTACCTTGGCTAA 1291
RESULT 6
AAT14037
ID AAT14037 standard; cDNA: 1269 BP.
XX
XX AAT14037;
XX AC
XX

DT	07-JUL-1996	(first entry)
XX	Transcription factor-IIIA gene.	
XX	Human: transcription factor-IIIA; hPRTIIA; DNA binding protein;	
KW	ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;	
KW	primer; PCR; polymerase chain reaction; foetal brain;	
KW	anchor primer; diagnostic; probe; transcription control;	
KW	antitumour; cancer; therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
EH	Key	Location/Qualifiers
FT	misc-feature	336..115
FT		/tag= a
FT		/product= Zinc finger domains
FT	polyA_signal	1382..1387
FT		/tag= b
XX		
PN	EP704526-A1.	
XX		
PD	03-APR-1996.	
XX		
PE	05-SEP-1995;	95EP-0113908.
XX		
PR	05-SEP-1994;	94JP-0211022.
XX		
PA	(SAKA) OTSUKA PHARM CO LTD.	
PI	Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;	
XX		
DR	WPI: 1996-173033/18.	
DR	P-PSDB; AAR91305.	
XX		
PT	Human Transcription Factor III A gene - useful in regulation of	
PT	transcription and for diagnosis and treatment of e.g. cancer related	
PT	diseases	
XX		
PS	Claim 2: Page 10-11; 17pp; English.	
XX		
CC	The sequence encodes human transcription factor-IIIA (hPRTIIA),	
CC	a DNA binding protein with 9 zinc finger domains, which is necessary	
CC	for the initiation of 5S RNA gene transcription, binding to an	
CC	internal control region of the 5S gene. A fuller cDNA sequence	
CC	with flanking regions is given in AAT14038. A fragment lacking a	
CC	5' portion of the gene has been isolated from a human foetal brain	
CC	cDNA library (ORF7-1), and the 5' portion of the gene has been	
CC	isolated by 5'-rapid amplification of cDNA ends using primers H11-R	
CC	(AAT14039), H11-E (AAT14040), H11-H (AAT14041) and Ap-2 (AAT14042),	
CC	complementary to anchor primer AAT14043. Reverse transcription using	
CC	H-11R is followed by anchor primer ligation and PCR using Ap-2 and	
CC	H11-E, to give a full-length cDNA, ORF7. The gene and its encoded	
CC	protein may be used in diagnosis, identification or therapy of	
CC	hereditary diseases such as cancer, or other diseases resulting from	
CC	abnormal transcriptional control, and to analyse the mechanisms	
CC	involved in their activity.	
XX		
XX		
SQ	Sequence 1269 BP: 358 A; 324 C; 329 G; 258 T; 0 other:	
Query Match	96.2%; Score 1167; DB 17; Length 1269;	
Best Local Similarity	99.3%; Pred. No. 4e-302;	
Matches 1203; Conservative	0; Mismatches 5; Indels 3; Gaps	3;
0Y	1 GTGGCGGGCGGGGGAAGGTTCACAGGAGACCGCTGGGCGGGCGCGCGGTTCGCCGCG	60
Db		
61	GTGGCGGGCGGGGGAAGGTTCACAGGAGACCGCTGGGCGGGCGGGCGCGGTTCGCCGCG	119
0Y	ACGTGTCTGTGGACAGTGGACGAGCGCCCTTGGCTTGGAGAGCGCGCGCCCTGGA	120
Db		
120	ACGTGTCTGTGGACAGTGGACGAGCGCCCTTGGCTTGGAGAGCGCGCGCCCTGGA	179
0Y	TCGCGCGGGCGGTGTGGTGTGTCTTGTGACATCGCCGAGCGCTTCAATTC	180

Db	130	TTCCGCGGCGGTGCTCCGCCAATGTCGGGTGTCTCTTGACCAATGCGCCAGCGCTTCATTTCG	239
Qy	181	AGCGGGGAGAGACTCACCTCCGACCCCGCGCGCCCGCGCTTCCAGAGAGGTTTCATCTG	240
Db	240	AGCGGGGAGAGGCTCAGCTCCGACCCCGCGCGCCCGCGCTTCCAGAGAGTTTCATCTG	299
Qy	241	CTTCTTCCCTGACCTGCAGCGCCCAATTACAGCAAAAGCCTGGAAGCTTGACGCGCACCTGTG	300
Db	300	CTTCTTCCCTGACTGACGCGCCCAATTACAGCAAAAGCCTGGAAGCTTGACGCGCACCTGTG	359
Qy	301	CAGGACACACGGGGGAGAGACCATTTTGTTGTGTACTATGAAGGGTGTGGCAAGGCTTCAT	360
Db	360	CAGGACACACGGGGGAGAGACCATTTTGTTGTGTACTATGAAGGGTGTGGCAAGGCTTCAT	419
Qy	421	TGCAAGCCACTGGGCTGTGATCAAAAATTCACACAAAATCAAACTTGAAGAAAACATTTTGA	480
Db	480	TGCAAGCCAAATGGTGTGATCAAAAATTCACACAAAATCAAACTTGAAGAAAACATTTTGA	539
Qy	421	TGCAAGCCACTGGGCTGTGATCAAAAATTCACACAAAATCAAACTTGAAGAAAACATTTTGA	480
Db	420	CAGGACTTACCATTTGAGCCGCCACATTCCTGACTCACACAGAGAAAAGCCGTTTGTG	479
Qy	361	CAGGACTTACCATTTGAGCCGCCACATTCCTGACTCACACAGAGAAAAGCCGTTTGTG	420
Db	361	CAGGACTTACCATTTGAGCCGCCACATTCCTGACTCACACAGAGAAAAGCCGTTTGTG	479
Qy	481	ACGCAAACTATAAATCAACAAAAACAAATATATGTCACTTTTGAAGACTGTAGAGAC	540
Db	540	ACGCAAACTATAAATCAACAAAAACAAATATATGTCACTTTTGAAGACTGTAGAGAAC	599
Qy	541	CTTTTAAAGAAATCATAGAGAGCTGAAATCCATCGTCCAGCATACCATAGAACCTCTATT	600
Db	600	CTTTTAAAGAAATCATAGAGAGCTGAAATCCATCGTCCAGCATACCATAGAACCTCTATT	659
Qy	601	CAAGTGTACCCAGAGAGAGATGTGGGAAACACTTTGGATACCCAGCAAGCTGAAGACGA	660
Db	660	CAAGTGTACCCAGAGAGAGATGTGGGAAACACTTTGGATACCCAGCAAGCTGAAGACGA	719
Qy	661	TGCCAAGGCCACGAGAGGCTATGTATGTCAAAAAGATGCTCTTTGTGGCAAAAACATG	720
Db	720	TGCCAAGGCCACGAGAGGCTATGTATGTCAAAAAGATGCTCTTTGTGGCAAAAACATG	779
Qy	721	GACGGAATTTGAAACATGTGAGAGAAACCCCTAAGAGGAAATATATGTGAAGTATG	780
Db	780	GACGGAATTTGAAACATGTGAGAGAAACCCCTAAGAGGAAATATATGTGAAGTATG	839
Qy	781	CCGGAACCATTTTAAAGCAAAAGATTACCTTAAGCAACATGAAACATCATGCCCCAGA	840
Db	840	CCGGAACCATTTTAAAGCAAAAGATTACCTTAAGCAACATGAAACATCATGCCCCAGA	899
Qy	841	AAGGATGTATGTGCTGTCCAGAGAGGCTGTGAGAACCTTACTACTGTGTTAA	900
Db	900	AAGGATGTATGTGCTGTCCAGAGAGGCTGTGAGAACCTTACTACTGTGTTAA	959
Qy	901	TCTCCAAAGCCATTCCTCTCTTCCATGAGGAAACCCGCTTTGTGTGTGAACATGC	960
Db	960	TCTCCAAAGCCATTCCTCTCTTCCATGAGGAAACCCGCTTTGTGTGTGAACATGC	1019
Qy	961	TGGCTGTGGCAAAACATTTGCATGAAACAAAGTCTACCTAGGCAATGCTTTGTACATGA	1020
Db	1020	TGGCTGTGGCAAAACATTTGCATGAAACAAAGTCTACCTAGGCAATGCTTTGTACATGA	1079
Qy	1021	TCTGTACAGAGAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAGAAAC -GGAGTTTGG	1079
Db	1080	TCTGTACAGAGAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAGAAACGGGAGTTTGG	1139
Qy	1080	CCTCTCAATCTAGTGAATATATCTCCCTCCAAAAGAAACAAAGGCAAGGCTTATCTTGT	1139
Db	1140	CCTCTCAATCTAGTGAATATAT -CTTCCCAAAAAGAAACAAAGGCAAGGCTTATCTTGT	1198
Qy	1140	GTCGAAAAGGAGATTCACCCCACTGTGTGGAAAGCAAGATGCTTCGACAGTTGCAGTAC	1199
Db	1199	GTCGAAAAGGAGATTCACCCCACTGTGTGGAAAGCAAGATGCTTCGACAGTTGCAGTAC	1258
Qy	1200	TTACCTTGGC 1210	
Db	1259	TTACCTTGGC 1269	

XX	AB054923/c	standard; cDNA; 690 BP.
XX	AB054923:	
XX	22-AUG-2002	(first entry)
DE	Human ovarian antigen HUSYA18 cDNA, SEQ ID NO:803.	
XX	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;	
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;	
KW	infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;	
KW	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;	
KW	inflammatory condition; immune disorder; blood disorder;	
KW	cardiovascular disorder; respiratory disorder; neurological disorder;	
KW	gastrointestinal disorder; urinary system disorder; drug screening;	
KW	gene therapy; chromosome mapping; forensic analysis;	
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;	
KW	antifertility; gynaecological; reproductive; chromosome 19q12.3-13.1;	
KW	gene; ss.	
OS	Homo sapiens.	
XX	WO200200677-A1.	
PN	03-JAN-2002.	
PD	07-JUN-2001; 2001MO-US18569.	
PE	07-JUN-2000; 2000US-209467P.	
PR	(HUMA-) HUMAN GENOME SCI INC.	
XX	Blirce CE, Rosen CA;	
PI	WPI: 2002-147878/19.	
DR	P-PSDB; ABPA1846.	
XX	Isolated nucleic acid molecules encoding novel ovarian polypeptides,	
PT	useful in the prevention, treatment and diagnosis of cancer (e.g.	
PT	ovarian cancer). Immune disorders, cardiovascular disorders and	
PT	neurological diseases -	
XX	Claim 1: SEQ ID NO 803; 2922pp; English.	
PS	The invention relates to 2175 novel human ovarian antigens (ABPA1054-	
CC	ABPA43228) and to cDNAs encoding them (AB054131-AB056305), and also	
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical	
CC	to the sequences of the invention. The invention additionally relates to	
CC	recombinant vectors and host cells comprising human ovarian antigen	
CC	polynucleotides, antibodies against human ovarian antigens, and the use	
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,	
CC	treating, prognosing or preventing various ovary and/or breast-related	
CC	disorders. Such conditions include ovarian cancer and breast cancer, and	
CC	metastatic tumours of ovarian or breast origin, reproductive system	
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,	
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine	
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic	
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and	
CC	vaginitis), immune disorders (e.g., congenital and acquired	
CC	immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus)	
CC	blood-related disorders (e.g., anaemia), cardiovascular disorders,	
CC	respiratory disorders, neurological disorders, gastrointestinal disorder	
CC	and urinary system disorders. Ovarian antigen polypeptides and	
CC	polynucleotides may also be used in screening for compounds which	
CC	modulate ovarian antigen expression or activity. The polynucleotides may	
CC	further be used for gene therapy, chromosome mapping, in the	
CC	identification of individuals and in forensic analysis, and the	
CC	polypeptides may be used as food additives or to prepare antibodies	
CC	useful in disease diagnosis, drug targeting and phenotyping. The present	

XX 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX
XX (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secretist H;
XX
DR WPI: 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1; SEQ ID No 326; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABR44450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 439 BP; 106 A; 94 C; 99 G; 140 T; 0 other;
XX
Query Match 34.5%; Score 418.4; DB 24; Length 439;
Best Local Similarity 99.8%; Pred. No. 6.4e-102;
Matches 419; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 794 AAGCGAAGATTACCTTAAGCAACATGAAACTGATGCCCGAAGAGGATGTATGT 853
DB 439 AAGCGAAGATTACCTTAAGCAACATGAAACTGATGCCCGAAGAGGATGTATGT 380
XX
QY 854 CGCTGTCCAGAGAGGCTGTGGAAGACCTATACCTGTGTATCTCCAAAGCCAT 913
DB 379 CGCTGTCCAGAGAGGCTGTGGAAGACCTATACCTGTGTATCTCCAAAGCCAT 320
XX
QY 914 ATCCCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGTGCAA 973
DB 319 ATCCCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGTGCAA 260
XX
QY 974 ACATTTCGAATGAACAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAGAG 1033
DB 259 ACATTTCGAATGAACAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAGAG 200
XX
QY 1034 AAAATGAAGCTCAAAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAG 1093
DB 199 AAAATGAAGCTCAAAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAG 140
XX
QY 1094 GGATATATCCCTCCCAAGAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAACGAGAG 1153
DB 139 GGATATATCCCTCCCAAGAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAACGAGAG 80
XX
QY 1154 TCACCAACTGTGTGAGACAGATGCTGTGACAGTGTGACATCTTACCTTGGCTTAA 1213
DB 79 TCACCAACTGTGTGAGACAGATGCTGTGACAGTGTGACATCTTACCTTGGCTTAA 20
XX
RESULT 9
ABK45138 ID ABR45138 standard; cDNA: 439 BP.
XX
AC ABR45138;
XX

DT 05-JUN-2002 (first entry)
XX
DE cDNA encoding colon tumour protein, SEQ ID No 689.
XX
KW Human: colon tumour; vaccine; colon cancer; immunogenic;
KW Immunotherapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212328-A2.
XX
PD 14-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US24218.
XX
PR 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX
PA (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secretist H;
XX
DR WPI: 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1; SEQ ID No 689; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABR44450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 439 BP; 141 A; 99 C; 93 G; 106 T; 0 other;
XX
Query Match 34.4%; Score 416.8; DB 24; Length 439;
Best Local Similarity 99.5%; Pred. No. 1.7e-101;
Matches 418; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
XX
QY 794 AAGCGAAGATTACCTTAAGCAACATGAAACTGATGCCCGAAGAGGATGTATGT 853
DB 1 AAGCGAAGATTACCTTAAGCAACATGAAACTGATGCCCGAAGAGGATGTATGT 60
XX
QY 854 CGCTGTCCAGAGAGGCTGTGGAAGACCTATACCTGTGTATCTCCAAAGCCAT 913
DB 61 CGCTGTCCAGAGAGGCTGTGGAAGACCTATACCTGTGTATCTCCAAAGCCAT 120
XX
QY 914 ATCCCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGTGCAA 973
DB 121 ATCCCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGTGCAA 180
XX
QY 974 ACATTTCGAATGAACAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAGAG 1033
DB 181 ACATTTCGAATGAACAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAGAG 240
XX
QY 1034 AAAATGAAGCTCAAAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAG 1093
DB 241 AAAATGAAGCTCAAAAGTCTCACTAGGCATGCTGTGTGATCATGATCTGACAAG 300
XX
QY 1094 GGATATATCCCTCCCAAGAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAACGAGAG 1153

|||||
Db 301 GGATATATCCCTCCCAAGAAAGGCGACAGCTTATCTTGTGCAAAACGAGAG 360
QY 1154 TCACCAACTGTGTGGAAGACATGCTCTCGACAGTGTGAGTACTTACCTGGCTAA 1213
Db 361 TCACCAACTGTGTGGAAGACATGCTCTCGACAGTGTGAGTACTTACCTGGCTAA 420
RESULT 10
ABO59499
ID ABO59499 standard; cDNA: 537 BP.
XX
AC ABO59499;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:3194.
XX
KM Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KM genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PE 02-OCT-2001: 2001MO-US0732.
XX
PR 02-OCT-2000: 2000US-237271P.
XX
PA (FARB) BAYER CORP.
XX
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thadlingam A, Lewis ME;
XX
DR WPI: 2002-426115/45.
XX
PT New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell
PT or tissue type, and in antisense therapy -
XX
PS Claim 1: Fig 1: 796pp; English.
XX
CC ABO56306 to ABO60787 represent isolated nucleic acids (1) differentially
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC encoded by the ABO60776 to ABO60787 nucleic acid sequences. (1) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (1) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (1) in a cell. A probe/primer derived
CC from (1) can be used for determining the presence of a nucleic acid which
CC hybridizes to (1), and for determining the phenotype of cells in a sample
CC of cells from a patient. (1) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC microarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (1) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists.
XX
SO Sequence 537 BP; 196 A; 116 C; 106 G; 117 T; 2 other:
Query Match 30.8%; Score 374; DB 24; Length 537;
Best Local Similarity 96.6%; Pred. No. 5.3e-90;
Matches 402; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
QY 602 AAGTGATCCCGAAGAGTGTGGAAACACTTGCATCACCAGCAAGCTGAACGACAT 661
Db 123 AAAAGGTCCCGAAGAGTGTGGAAACACTTGCATCACCAGCAAGCTGAACGACAT 182
QY 662 GCCAAGGCCGACGAGCGCTATGTATGTCAAAAGATGCTTGTGCAAAAACATGG 721
|||||

Db 183 GCCAAGGCCGACGAGCGCTATGTATGTCAAAAGATGCTTGTGCAAAAACATGG 242
QY 722 ACGGACTTCGAAACATGTGAGAGAAACCCATGAAGGAAATATGATGAGATGTC 781
Db 243 ACGGACTTCGAAACATGTGAGAGAAACCCATGAAGGAAATATGATGAGATGTC 302
QY 782 CGGAAACATTTAAACGCAAGATTTACCTTAAGCAACATGAAACATGATGCCCCGAA 841
Db 303 CGGAAACATTTAAACGCAAGATTTACCTTAAGCAACATGAAACATGATGCCCCGAA 362
QY 842 AGGATGTATGTGCTGTGCAAGAGAGCTGTGGAAGACCTATACTGTGTTTAT 901
Db 363 AGGATGTATGTGCTGTGCAAGAGAGCTGTGGAAGACCTAT -CAACTGTGTTTAT 421
QY 902 CTCGAAAGCCATATCTCTCTCTCCATGAGGAAAGCCGCTTTGTGTGAAACATGCT 961
Db 422 CTCGAAAGCCATATCTCTCTCCATGAGGAAAGCCGCTTTGTGTGAAACATGCT 481
QY 962 GCGTGTGCAAAACA -TTTGCAATGAACAAGTCTCAGTGCATGCTGTGTAC 1016
Db 482 GCGTGTGCAAAACATTTTGCATGAAACATGCTCAGTGCATGCTGTGTAC 537
RESULT 11
ABV49231
ID ABV49231 standard; cDNA: 466 BP.
XX
AC ABV49231;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 49222.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001: 2001MO-US05171.
XX
PR 17-FEB-2000: 2000US-183319P.
PR 16-MAR-2000: 2000US-189862P.
PR 25-MAY-2000: 2000US-207454P.
PR 09-JUN-2000: 2000US-211314P.
PR 18-JUL-2000: 2000US-219007P.
PR 13-DEC-2000: 2000US-255281P.
XX
PA (MILL-) MILENITUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1: Page 9624: 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 466 BP; 153 A; 92 C; 102 G; 119 T; 0 other;
Query Match 24.8%; Score 301.4; DB 23; Length 466;
Best Local Similarity 99.7%; Pred. No. 1.4e-70;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 677 GCGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGAGGAACTTGTGAAA 736
DB 164 GCGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGAGGAACTTGTGAAA 223
OY 737 CATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATCCGGAAAACTTTAAA 796
DB 224 CATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATCCGGAAAACTTTAAA 283
OY 797 CGCAAAAGATTACTTTAAGCAACACATGAAACATGAGGAAAGGATGTATGTCGC 856
DB 284 CGCAAAAGATTACTTTAAGCAACACATGAAACATGAGGAAAGGATGTATGTCGC 343
OY 857 TGTCCAAAGAGAGGCTGTGGAAGAACTATACACTGTGTTTAATCTCCAAAGCCATATC 916
DB 344 TGTCCAAAGAGAGGCTGTGGAAGAACTATACACTGTGTTTAATCTCCAAAGCCATATC 403
OY 917 CTCTCCTTCCATGAGGAAAGCCGCTTTTGTGTGAACATGCTGGCTGTGCAAAAACA 976
DB 404 CTCTCCTTCCATGAGGAAAGCCGCTTTTGTGTGAACATGCTGGCTGTGCAAAAACA 463
OY 977 TTT 979
DB 464 TTT 466
RESULT 12
ABV19459
ID ABV19459 standard; cDNA; 425 BP.
XX
AC ABV19459;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 19450.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
XX
PR 25-MAY-2000; 2000US-207454P.
XX
PR 09-JUN-2000; 2000US-211314P.
XX
PR 18-JUL-2000; 2000US-219007P.
XX
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI: 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 3184; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 425 BP; 142 A; 79 C; 90 G; 114 T; 0 other;
Query Match 24.8%; Score 300.4; DB 23; Length 425;
Best Local Similarity 99.7%; Pred. No. 2.5e-70;
Matches 301; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 677 GCGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGAGGAACTTGTGAAA 736
DB 124 GCGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGAGGAACTTGTGAAA 183
OY 737 CATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATCCGGAAAACTTTAAA 796
DB 184 CATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATCCGGAAAACTTTAAA 243
OY 797 CGCAAAAGATTACTTTAAGCAACACATGAAACATGAGGAAAGGATGTATGTCGC 856
DB 244 CGCAAAAGATTACTTTAAGCAACACATGAAACATGAGGAAAGGATGTATGTCGC 303
OY 857 TGTCCAAAGAGAGGCTGTGGAAGAACTATACACTGTGTTTAATCTCCAAAGCCATATC 916
DB 304 TGTCCAAAGAGAGGCTGTGGAAGAACTATACACTGTGTTTAATCTCCAAAGCCATATC 363
OY 917 CTCTCCTTCCATGAGGAAAGCCGCTTTTGTGTGAACATGCTGGCTGTGCAAAAACA 976
DB 364 CTCTCCTTCCATGAGGAAAGCCGCTTTTGTGTGAACATGCTGGCTGTGCAAAAACA 423
OY 977 TT 978
DB 424 TT 425
RESULT 13
AAH22129/c
ID AAH22129 standard; cDNA; 1401 BP.
XX
AC AAH22129;
XX
DT 20-AUG-2001 (first entry)
XX
DE Human eukaryotic initiation factor 3 (heIF3) encoding cDNA SEQ ID NO:6.
XX
DE Human eukaryotic initiation factor 3 (heIF3) encoding cDNA SEQ ID NO:6.
XX
KW Human; eIF3; eukaryotic initiation factor; translation initiation factor;
KM heIF3; dendritic cell; ss.
XX
XX Homo sapiens.
XX
OS
XX
XX Key Location/Qualifiers
XX FT 237..1073
XX FT CDS /*tag=a
XX FT /product="eukaryotic initiation factor (heIF3)"
XX
XX CNI281044-A.

PD 24-JAN-2001.
XX
XX 27-JUN-2000; 2000CN-0116791.
PF
XX 27-JUN-2000; 2000CN-0116791.
PR
XX
XX (NAME-) NANFANG RES CENT STATE HUMAN GENE GROUP.
PA
XX
XX Yang Y, Xiao H, Kang B:
PI
XX WPI; 2001-282654/30.
DR
XX P-PSDB: AAB98214.
DR
XX
XX Human translation initiation factor protein and its coding sequence -
PT
XX
XX
PS Claim 1; Page 17 (disclosure); 20pp; Chinese.
XX
XX The present invention describes a human eukaryotic initiation factor
CC (eIF), designated heif3, which is expressed in human dendritic cells.
CC heif3 has translation initiation factor activity. Also described in the
CC present invention are methods for the preparation and detection of the
CC heif3 protein and nucleotide sequences. The present sequence encodes
CC heif3, as given in the present invention.
XX
XX Sequence 1401 BP; 459 A; 281 C; 306 G; 355 T; 0 other;
SQ
Query Match 20.3%; Score 246; DB 22; Length 1401;
Best Local Similarity 100.0%; Pred. No. 1.6e-55;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 968 GGGCAAAACATTTGCAATGAAACAAAGTCACATGAGCATGCTGTTGATGATCCTGAC 1027
DB 1401 GGGCAAAACATTTGCAATGAAACAAAGTCACATGAGCATGCTGTTGATGATCCTGAC 1342
QY 1028 AACAGAAATGAAAGCTCAAAAGTCAAAATTCGCTGAAAAACGAGATTGGCCTCTCAT 1087
DB 1341 AACAGAAATGAAAGCTCAAAAGTCAAAATTCGCTGAAAAACGAGATTGGCCTCTCAT 1282
QY 1088 CTGAGTGAATATTCCTCCCAAGGAAGCAAGGCGATATCTTGTGTCAAAC 1147
DB 1281 CTGAGTGAATATTCCTCCCAAGGAAGCAAGGCGATATCTTGTGTCAAAC 1222
QY 1148 GGAAGATCACCACTGTGTGGAAGACAGATGCTCTCGACAGTGGAGTACTTACCCTT 1207
DB 1221 GGAAGATCACCACTGTGTGGAAGACAGATGCTCTCGACAGTGGAGTACTTACCCTT 1162
QY 1208 GCGCTAA 1213
DB 1161 GCGCTAA 1156
RESULT 14
AAS73534
ID AAS73534 standard; cDNA: 1889 BP.
AC AAS73534;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #9338.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200175067-A2.
PN
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR

PR 23-AUG-2000; 2000US-0649167.
PA
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB: ABG09347.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX
PS Claim 1; SEQ ID No 9338; 103pp; English.
XX
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
SQ Sequence 1889 BP; 488 A; 453 C; 417 G; 531 T; 0 other;
Query Match 19.5%; Score 236; DB 23; Length 1889;
Best Local Similarity 95.5%; Pred. No. 9e-53;
Matches 275; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
QY 926 CATGAGAAACCGCCCTTTGTGTGGAACATGCGTGGCAAAACATTGCAATG 985
DB 2 CATGAGAAACCGCCCTTTGTGTGGAACATG-TGGCTGTGGCAAAACA-TTGCAAAT 59
QY 986 AACCAAAAGTCTCACTAGGATGCTGTTGTACATGATCTGACAAAGAAATGAAGCTC 1045
DB 60 AACCAAAAGTCTCACTAGGATGCTGTTGTACATG-TGCTGTGCAAAAGAAATGAAGCTC 118
QY 1046 AAGTCAAAATCTGCTGAAAAAAGGAGTTGGCTCTCATCTCACTGATATATCCCT 1105
DB 119 AAGTCAAAATCTGCTGAAAAAAGGAGTTGGCTCTCATCTCACTGATATATCCCT 178
QY 1106 CCCAAAAGAAACAAAGGCGATATCTTGTGTCAAAAGGAGAGTACCACTG 1165
DB 179 CCCAAAAGAAACAAAGGCGATATCTTGTGTCAAAAGGAGAGTACCACTG 238
QY 1166 GTGGAAGACAAGATGCTCTGACAGATTGACGATCTTACCCTTGGCTAA 1213
DB 239 GTGGAAGACAAGATGCTCTGACAGATTGACGATCTTACCCTTGGCTAA 286
RESULT 15
ABK45059
ID ABK45059 standard; cDNA: 234 BP.
AC ABK45059;
XX
XX
XX 05-JUN-2002 (first entry)
DT
XX

DE cDNA encoding colon tumour protein, SEQ ID No 610.
XX
KM Human; colon tumour; vaccine; colon cancer; immunogenic;
KW Immunotherapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212328-A2.
XX
PD 14-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US24218.
XX
PR 03-AUG-2000; 2000US-223283P.
XX
PR 28-MAR-2001; 2001US-279763P.
XX
PR 29-JUN-2001; 2001US-302051P.
XX
PA (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secretist H;
XX
DR WPI; 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
XX for diagnosing, preventing, and treating colon cancer, and as markers
XX for the progression of cancer -
XX
PS Claim 1; SEQ ID No 610; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
XX The polynucleotides and encoded polypeptides are useful in pharmaceutical
XX compositions, such as vaccines, for the diagnosis, prevention, and
XX treatment of colon cancer. Polynucleotide sequences may be used as
XX hybridisation probes or primers, and in the design and preparation of
XX ribozyme molecules for inhibiting expression of tumour polypeptides and
XX proteins in tumour cells. The compositions are useful for stimulating an
XX immune response against cancer, particularly for the immunotherapy of
XX colon cancer, and as markers for the progression of cancer.
XX ABK4450-ABK46237 represent coding sequences of human colon tumour
XX proteins of the invention.
XX Note: With the exception of SEQ ID No 1 and 2, the sequence data
XX for this patent did not form part of the printed specification but was
XX supplied by the European Patent Office.
XX
SQ Sequence 234 Bp: 83 A; 50 C; 53 G; 48 T; 0 other;

Query Match 19.3%; Score 234; DB 24; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.1e-52;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 CTGAATCCATCGATGCGCAGCATACCAATGACCTCTATCAAGTGTACCCAGGAAGA 619
DB 1 CTGAATCCATCGATGCGCAGCATACCAATGACCTCTATCAAGTGTACCCAGGAAGA 60
QY 620 TGTGGAAACACTTGTGCTACCCAGCAAGCTGAACGACATGCCAAGGCCACAGAGGC 679
DB 61 TGTGGAAACACTTGTGCTACCCAGCAAGCTGAACGACATGCCAAGGCCACAGAGGC 120
QY 680 TATGATGTCAAAAAGATGTTCTTGTGTGCAAAAACATGACGCGAATCTCTGAACAT 739
DB 121 TATGATGTCAAAAAGATGTTCTTGTGTGCAAAAACATGACGCGAATCTCTGAACAT 180
QY 740 GTGAGAGAAACCATTAAGAGAAATACTATGTAAGTATGCCGAAACATTT 793
DB 181 GTGAGAGAAACCATTAAGAGAAATACTATGTAAGTATGCCGAAACATTT 234

Search completed: February 10, 2003, 12:56:26
Job time : 293.158 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 13:06:36 : Search time 21 Seconds

(without alignments)
720.898 Million cell updates/sec

Title: US-09-831-426C-2

Perfect score: 1 MDPVAVAESVSLTIADAF.....ESPNCVEDKMLSTAVAVLTIG 365

Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%

Listing first 45 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1993	99.3	423	TF3A_HUMAN	Q92664 homo sapien
2	1055.5	52.6	366	TF3A_XENLA	P03001 xenopus lae
3	1047	52.1	335	TF3A_RANPI	P34695 rana pipien
4	1024	51.0	339	TF3A_XENBO	P17842 xenopus bor
5	1022	50.9	339	TF3A_BUFAM	P34694 bufu amicic
6	816	40.6	322	TF3A_ICTPU	P79797 ictalurus p
7	556.5	27.7	365	P43_XENLA	P23436 xenopus lae
8	553.5	27.6	365	P43_XENBO	P23066 xenopus bor
9	417.5	20.8	803	ZXDB_HUMAN	P98169 homo sapien
10	410.5	20.4	616	ZN93_HUMAN	P35789 homo sapien
11	409	20.4	336	ZG57_XENLA	P18729 xenopus lae
12	409	20.4	799	ZKRA_HUMAN	P98168 homo sapien
13	406	20.2	698	Z234_HUMAN	O14588 homo sapien
14	404	20.1	821	ZN41_HUMAN	P51814 homo sapien
15	403	20.1	574	YE73_HUMAN	O98255 homo sapien
16	399	19.9	1167	Z208_HUMAN	O43345 homo sapien
17	398.5	19.8	626	Z143_HUMAN	P52747 homo sapien
18	398.5	19.8	751	Z337_HUMAN	O973m9 homo sapien
19	395.5	19.7	453	Z06_XENLA	P18749 xenopus lae
20	395	19.6	803	Z226_HUMAN	O99yt6 homo sapien
21	394.5	19.6	803	ZN43_HUMAN	P17038 homo sapien
22	393	19.6	689	EZIT_HUMAN	O98qz8 homo sapien
23	393	19.6	689	PRD9_HUMAN	O98qv7 homo sapien
24	392.5	19.5	337	ZN81_HUMAN	P51508 homo sapien
25	391	19.5	738	ZN84_HUMAN	P51523 homo sapien
26	386.5	19.2	759	Z287_MOUSE	O9eqb9 mus musculu
27	386.5	19.2	898	Z071_XENLA	P18751 xenopus lae
28	386	19.2	913	Z228_HUMAN	O99tj3 homo sapien
29	385	19.2	626	Z189_HUMAN	O75820 homo sapien
30	383	19.1	538	Z155_HUMAN	O12801 homo sapien
31	382.5	19.0	280	ZG46_XENLA	P18722 xenopus lae
32	382	19.0	645	ZF93_MOUSE	O61116 mus musculu
33	381.5	19.0	535	Z257_HUMAN	O9y2q1 homo sapien

34	381.5	19.0	734	ZN42_HUMAN	P28698 homo sapien
35	381	19.0	711	Z175_HUMAN	O9y473 homo sapien
36	380.5	18.9	595	ZN85_HUMAN	O03923 homo sapien
37	380.5	18.9	682	ZN45_HUMAN	O02386 homo sapien
38	379	18.9	683	Y972_HUMAN	O9y2h8 homo sapien
39	378	18.8	652	Y798_HUMAN	O94892 homo sapien
40	377.5	18.8	595	Z317_HUMAN	O96p96 homo sapien
41	377	18.8	604	Z300_HUMAN	O96p96 homo sapien
42	376.5	18.8	446	ZN70_HUMAN	O9uc06 homo sapien
43	376.5	18.8	472	Z1M3_HUMAN	O96p66 homo sapien
44	376.5	18.8	570	ZN76_HUMAN	P36508 homo sapien
45	376.5	18.8	697	HKR1_HUMAN	P10072 homo sapien

ALIGNMENTS

```

RESULT 1
ID TF3A_HUMAN STANDARD: PRT: 423 AA.
AC Q92664: Q13097: Q12963:
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor IIA (Factor A) (TFIIIA).
GN TF3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95309028; PubMed=7789179;
RA Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,
RA Fujiwara T., Takahashi E., Shin S., Nakamura Y.;
RT "Molecular cloning, characterization, and chromosomal mapping of a
RT novel human gene (GTF3A) that is highly homologous to Xenopus
RT transcription factor IIA."
RT Cytogenet. Cell Genet. 70:235-238(1995).
RN [2]
RP SEQUENCE OF 61-423 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95347600; PubMed=7622052;
RA Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.,
RA Becker K.G.;
RT "Cloning and expression analysis of a human cDNA homologous to
RT Xenopus TFIIIA."
RT Gene 159:215-218(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=94342241; PubMed=8063702;
RA Moorefield B., Koeder R.G.;
RT "Purification and characterization of human transcription factor
RT IIA."
RT J. Biol. Chem. 269:20857-20865(1994).
RN [4]
RP FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
RP APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR
RP CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
RP BINDS THE TRANSCRIBED 5S RNA'S. MAY INITIATE TRANSCRIPTION OF THE
RP 5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION
RP OF OTHER GENES.
RN [5]
RP SUBCELLULAR LOCATION: Nuclear.
RN [6]
RP TISSUE SPECIFICITY: UBICUITOUS.
RN [7]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb.ch/announce/
RP or send an email to license@isb-sib.ch).

```

DR EMBL: D32257; BAA06988.1; -
 DR EMBL: U20277; AAA75623.1; -
 DR EMBL: U14134; AAA21873.1; -
 DR HSSP: P03001; 1TF3.
 DR TRANSFAC: T04953; -
 DR Genew: HGNC:4662; GTF3A.
 DR MIM: 600860; -
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; ZF-C2H2; 9.
 DR PRINTS: PRO0048; ZINC_FINGER.
 DR SMART: SM00355; Znf_C2H2; 8.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
 DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 DR RNA-binding; Repeat; Nuclear protein.
 FT DOMAIN 98 359 ZINC_FINGERS.
 FT ZN_FING 98 122 C2H2-TYPE.
 FT ZN_FING 128 152 C2H2-TYPE.
 FT ZN_FING 158 183 C2H2-TYPE.
 FT ZN_FING 190 212 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 220 244 C2H2-TYPE.
 FT ZN_FING 247 271 C2H2-TYPE.
 FT ZN_FING 275 297 C2H2-TYPE.
 FT ZN_FING 304 329 C2H2-TYPE.
 FT ZN_FING 335 359 C2H2-TYPE.
 FT CONFLICT 214 214 N -> H (IN REF. 2).
 FT CONFLICT 221 221 K -> KL (IN REF. 2; AAA21873).
 FT CONFLICT 223 248 MISSING (IN REF. 2; AAA21873).
 FT CONFLICT 378 387 SLASHSGYT -> EFGSSOMY (IN REF. 1).
 SQ SEQUENCE 423 AA: 46847 MW: A627D064A43FE6F0 CR664.

Query Match 99.3%; Score 1993; DB 1; Length 423;
 Best Local Similarity 99.2%; Pred. No. 8, 5e-146;
 Matches 362; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 MPPAVAVASVSLTIADFIAGESSATPPRPALPRRFICSPFDCSANYSKAKLDAH 60
 :|||||
 DB 59 LPPAVAVASVSLTIADFIAGESSATPPRPALPRRFICSPDCSANYSKAKLDAH 118
 OY 61 LCKHTGERFVCDYEGCGAFIRDYHLRHLTHGGERFVCAAGCDOKFMTKSNLKKH 120
 :|||||
 DB 119 LCKHTGERFVCDYEGCGAFIRDYHLRHLTHGGERFVCAAGCDOKFMTKSNLKKH 178
 OY 121 FERKHNQKQYICSEFDCCKTFKKHQQLKHCQHTNPLFKCTQEGCGKHPASPKLK 180
 :|||||
 DB 179 FERKHNQKQYICSEFDCCKTFKKHQQLKHCQHTNPLFKCTQEGCGKHPASPKLK 238
 OY 181 RHAKAHEGVCCGCSFVAKTTELLKHYRETHKEELCEVCRKTFKRKYLOKHKHTHA 240
 :|||||
 DB 239 RHAKAHEGVCCGCSFVAKTTELLKHYRETHKEELCEVCRKTFKRKYLOKHKHTHA 298
 OY 241 PERDVCRCRPGCGRTYTVFNLOSHILSFHESRPFCVEHAGCGTFAMKOSLTPHAVV 300
 :|||||
 DB 299 PERDVCRCRPGCGRTYTVFNLOSHILSFHESRPFCVEHAGCGTFAMKOSLTPHAVV 358
 OY 301 HDPDKKKMKLVKKSRKESLASLSGVIIPRKQOGSLCQNGESPNCVEDKMLSTVA 360
 :|||||
 DB 359 HDPDKKKMKLVKKSRKESLASLSGVIIPRKQOGSLCQNGESPNCVEDKMLSTVA 418
 OY 361 VLTIG 365
 :|||||
 DB 419 VLTIG 423

RESULT 2

TF3A_XENLA STANDARD: PRT: 366 AA.
 ID TF3A_XENLA
 AC P03001; Q91856;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor Iiia (Factor A) (TFIIIA) (S-TFIIIA/O-TFIIIA).
 OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE OF 23-366 FROM N.A.
 RX MEDLINE=85074456; PubMed=6210149;
 RA Ginsberg A.M., King B.O., Roeder R.G.;
 RT "Xenopus 5S gene transcription factor, TFIIIA: characterization of a
 RT cDNA clone and measurement of RNA levels throughout development.";
 RL Cell 39:479-489(1984).
 RN [2]
 RP SEQUENCE OF 23-366 FROM N.A.
 RX MEDLINE=86176722; PubMed=3754326;
 RA Yun Tso J., van den Berg J., Korn L.J.;
 RT "Structure of the gene for Xenopus transcription factor TFIIIA";
 RL Nucleic Acids Res. 14:2187-2201(1986).
 RN [3]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 112-137.
 RX MEDLINE=86312888; PubMed=3755818;
 RA Taylor M., Jackson I.J., Siegel N., Kumar A., Brown D.D.;
 RT "The developmental expression of the gene for TFIIIA in Xenopus
 RT laevis";
 RL Nucleic Acids Res. 14:6185-6195(1986).
 RN [4]
 RP SEQUENCE OF 1-35 FROM N.A.
 RX MEDLINE=89306603; PubMed=2744458;
 RA Scotti K.W., Kaulen H., Roeder R.G.;
 RT "Positive and negative regulation of the gene for transcription
 RT factor Iiia in Xenopus laevis oocytes";
 RL Genes Dev. 3:651-662(1989).
 RN [5]
 RP SEQUENCE OF 1-26 FROM N.A.
 RX MEDLINE=91071587; PubMed=2253880;
 RA Kim S.H., Dady M.K., Joho K.E., Brown D.D.;
 RT "The characterization of the TFIIIA synthesized in somatic cells of
 RT Xenopus laevis";
 RL Genes Dev. 4:1602-1610(1990).
 RN [6]
 RP REPEATS ANALYSIS.
 RX MEDLINE=85284956; PubMed=4040853;
 RA Miller J., McLachlan A.D., Kling A.;
 RT "Repetitive zinc-binding domains in the protein transcription factor
 RT Iiia from Xenopus oocytes";
 RL EMBO J. 4:1609-1614(1985).
 RN [7]
 RP REPEATS ANALYSIS.
 RX MEDLINE=85231235; PubMed=4007166;
 RA Brown R.S., Sander C., Argos P.;
 RT "The primary structure of transcription factor TFIIIA has 12
 RT consecutive repeats";
 RL FEBS Lett. 186:271-274(1985).
 RN [8]
 RP REPEATS ANALYSIS.
 RA Boehm S., Drescher B.;
 RT "Multiple internal repeats within the structure of the 5S RNA/DNA
 RT binding transcription factor TF-Iiia from Xenopus laevis";
 RL Studia Biophys. 107:237-247(1985).
 RN [9]
 RP STRUCTURE BY NMR OF 10-101.
 RX MEDLINE=97397344; PubMed=9253405;
 RA Foster M.P., Muttke D.S., Radhakrishnan I., Case D.A.,
 RA Gottesfeld J.M., Wright P.E.;
 RT "Domain packing and dynamics in the DNA complex of the N-terminal
 RT zinc fingers of TFIIIA";
 RL Nat. Struct. Biol. 4:605-608(1997).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 32-210.
 RX MEDLINE=98169475; PubMed=9501194;
 RA Nolte R.T., Conlin R.M., Harrison S.C., Brown R.S.;
 RT "Differing roles for zinc fingers in DNA recognition: structure of a
 RT six-finger transcription factor Iiia complex";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:2938-2943(1998).

CC	-1	FUNCTION: ACTS AS BOTH A POSITIVE TRANSCRIPTION FACTOR FOR 5S RNA GENES AND A SPECIFIC RNA BINDING PROTEIN THAT COMPLEXES WITH 5S RNA IN OOCYTES TO FORM THE 7S RIBONUCLEOPROTEIN STORAGE PARTICLE. MAY PLAY AN ESSENTIAL ROLE IN THE DEVELOPMENTAL CHANGE IN 5S RNA GENE EXPRESSION. INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO BINDS THE TRANSCRIBED 5S RNA'S.	
CC	-1	SUBCELLULAR LOCATION: Nucleat.	
CC	-1	ALTERNATIVE PRODUCTS: 2 ISOFORMS; 5'-UTR (IN SOMATIC CELLS) AND 0'-UTR (IN OOCYTES); ARE PRODUCED BY ALTERNATIVE INITIATION.	
CC	-1	TISSUE SPECIFICITY: SYNTHESIZED IN OOCYTES AND, IN MUCH LOWER LEVELS, IN SOMATIC CELLS.	
CC	-1	DEVELOPMENTAL STAGE: THE LEVELS FOLLOW THE TRANSCRIPTIONAL ACTIVITY OF OOCYTE TYPE 5S RNA GENES DURING EMBRYOGENESIS, PRESENT IN VERY HIGH LEVELS IN MATURING OOCYTES WHEN OOCYTE TYPE 5S GENES ARE BEING EXPRESSED, AND IN MUCH LOWER LEVELS IN SOMATIC CELLS WHERE THE OOCYTE TYPE GENES ARE NOT EXPRESSED.	
CC	-1	PTM: THE N-TERMINUS IS BLOCKED.	
CC		-----	
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC		-----	
DR	EMBL	K02938; AAA49967.1; -	
DR	EMBL	X03681; CAB51745.1; -	
DR	EMBL	X03735; CAB51745.1; JOINED.	
DR	EMBL	X03736; CAB51745.1; JOINED.	
DR	EMBL	X03737; CAB51745.1; JOINED.	
DR	EMBL	X03738; CAB51745.1; JOINED.	
DR	EMBL	X03739; CAB51745.1; JOINED.	
DR	EMBL	X15785; CAB3786.1; -	
DR	PIR	A03529; TMXL3.	
DR	PIR	A24961; A24961.	
DR	PIR	S40785; S40785.	
DR	PDB	1TF3; 17-SEP-97.	
DR	PDB	1TF6; 08-JUL-98.	
DR	TRANSFAC	T02295; -	
DR	InterPro	IPR000822; Znf_C2H2.	
DR	Pfam	PF00096; Zf-C2H2; 8.	
DR	PRINTS	PRO0048; ZINCINGER.	
DR	SMART	SM00355; Znf_C2H2; 9.	
DR	PROSITE	PS00028; ZINC_FINGER_C2H2_1; 8.	
DR	PROSITE	PS50157; ZINC_FINGER_C2H2_2; 7.	
KM		Transcription regulation; Zinc-finger; Metal-binding; DNA-binding; RNA-binding; Repeat; Nuclear protein; Alternative initiation;	
KW		Polymorphism; 3d-structure.	
FT	CHAIN	1 366	TRANSCRIPTION FACTOR IIVA, SOMATIC ISOFORM.
FT	CHAIN	23 366	TRANSCRIPTION FACTOR IIVA, OOCYTE ISOFORM.
FT	INIT_MET	23 23	FOR OOCYTE ISOFORM.
FT	DOMAIN	35 298	ZINC FINGERS.
FT	ZN_FING	35 59	C2H2-TYPE.
FT	ZN_FING	65 89	C2H2-TYPE.
FT	ZN_FING	95 120	C2H2-TYPE.
FT	ZN_FING	127 151	C2H2-TYPE.
FT	ZN_FING	157 181	C2H2-TYPE.
FT	ZN_FING	184 210	C2H2-TYPE.
FT	ZN_FING	214 236	C2H2-TYPE.
FT	ZN_FING	243 268	C2H2-TYPE.
FT	ZN_FING	274 298	C2H2-TYPE.
FT	VARIANT	96 96	T -> K.
FT	VARIANT	319 319	R -> C.
FT	VARIANT	335 335	G -> I.
FT	VARIANT	356 356	G -> D.
FT	VARIANT	365 365	I -> L.
FT	CONFLICT	4 4	K -> T (IN REF. 4).
SO	SEQUENCE	366 AA; 41911 MW; 29237E710292FA24 CRC64;	

```

Query Match          52.6%; Score 1055.5; DB 1; Length 366;
Best Local Similarity 55.1%; Pred. No. 5.3e-74;
Matches 196; Conservative 46; Mismatches 107; Indels 7; Gaps 4;

QY 5 AVVAESVSLTTIADAFIAGESSAPTPRPRLPPRFICISFPKCSANYSKAMKLDALSKH 64
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 2 AAKVASTSESEAEGLSVLVEEMGEKA--LPVYKRYKICSPADCCAGAAKMKMLQANLCKH 59
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 65 TGERPFVCDYBSCGAFITRDYHLSHLITPTNGEKFVCAALGSCDOKETNTKSLMKKHFERK 124
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 60 TGEKPFPEKKEBCECEGKGFSLNHLTKHSLHNGEKNFTDSDGCDLRFPTKANKMKKHFRF 119
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 125 HENQOKVICSEPCDKKTFKKHQOKINOCOSTNPRLPKCEQEGSGKFPASPKLRKAK 184
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 120 HNIKLCVYVCHFEKNGKAFKKNQJAKVQFHSITQOLPPECSEHGSCDKRFSLPSRLKREK 179
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 185 AHGIVYCKR--GCSFVAATWTTELLKHNVEETHKEELICEVCRKTFKKRKYDLKOHMTAPE 242
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 180 VHAGYPCKKDDSCSPVGTWTTLYLKHNVAECHOQDLADVCNKKRFGNKHQLRDHOKTHKE 239
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 243 RDVYCPREGCGRTTYTPNLOSILSHEDSRPVYCNHAGSCGKTFRAMKOSITRAVYVD 302
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 240 RTVYLCPRPDGDRSYTTFNLSHISQSHEDRPVCEHAGCGKCFAMKKSLSRHSHVND 299
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 303 PDKKKMKLVKKRSRKRSLSHLSGTYTPKRRKQGGSLGNGESPNVCEOKMLST 358
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 300 PEKR--KLKEKCPRRKRSLSLMLTYGIRPKKE-KNAYSGETEKIDSLYKKNPKST 352
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 3
TF3A-RANP1
ID TF3A-RANP1 STANDARD: PRT: 335 AA.
AC P34695.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Transcription factor IITA (Factor A) (TFIIIA).
OS Rana pipiens (Northern leopard frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=8404;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE=93013035; PubMed=1398134;
RA Gaskins C.J., Smith J.F., Ogilvie M.K., Hannas J.S.;
RT "Comparison of the sequence and structure of transcription factor
RT IITA from Bufo americanus and Rana pipiens.";
RL Gene 120.197-206(1992).
CC -!- FUNCTION: INTERACTS WITH THE INTRNL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA S.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@sib-sib.ch).
CC -----
Cc EMBL: M8511; AAA49534.1; -.
Dr EMBL: X58369; CAA41260.1; -.
Dr PIR: JCI441; JCI441.
Dr HSSP: P03001; ITP3.
Dr InterPro: IPR000822; ZnF_C2H2.
Dr Pfam: PF00096; Zf-C2H2; 9
Dr PRINTS: PR00048; ZINCFINGER.
Dr SMART: SM00355; ZnF_C2H2; 9.
Dr PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.

```

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW RNA-binding; Repeat; Nuclear protein.
FT DOMAIN 13 276
FT ZN_FING 13 37
FT ZN_FING 43 67
FT ZN_FING 73 98
FT ZN_FING 105 129
FT ZN_FING 135 159
FT ZN_FING 162 188
FT ZN_FING 192 214
FT ZN_FING 221 246
FT ZN_FING 252 276
SQ SEQUENCE 335 AA; 38168 MM; E0B89EFD201DD288 CRC64;

Query Match 52.1%; Score 1047; DB 1; Length 335;
Best Local Similarity 62.1%; Pred. No. 2,2e-73;
Matches 187; Conservative 32; Mismatches 76; Indels 6; Gaps 2

OY	34	PALPREFTCSPEDCSGANSKAWKLDNALHNHGERFVCDYEGCGKAFIRDYHLNSHILT	93
Db	7	PAYVKRYTICSFADCSASYKNMKLQDLAHNLCKNHGERFPCTVEGCGKGFYTLFLRLHSMT	66
OY	94	HTEGEFPCATAGCCOQKEFKTSNLKKHFPRKDNDQOKYTICSFEDECKTKFFKHQOLKITQ	153
Db	67	HNGEPKCCDAPDCDLSEFTWTNLRLRHHYORAILSPLEICYFADCGGTFRKNNQKLHQ	126
OY	154	CQHTNEPLFKTCQECCGKGHPASPKLRRAKHEGVCOK--GCSEFVAITWTFELKHIVE	211
Db	127	YHTNQOPRKCHEGCDSFSFPSRLKRHEKYHACGRPOKDDSCSRVGCTWRPYMKNILA	186
OY	212	THKEILCEVCARKTRKRDYLKOHMKTAPEVDVCRCPREGCGRTYTVFNLSHLSIFR	271
Db	187	SHSEPTEICVNCRRKRNKTNHLLDKDHKRTHEVERVVYKCRPDGCDRTYTKRFGLSHLSIFR	246
OY	272	EHSRPVFCBHACCSTPTFAWKOSLTRAAVVDHDDKKMKLKLVKRSREKBSLASLSGYTPR	331
Db	247	EDSRFPACGHPCGKTFAWKOSLDRIANTHDPEKKM---KKPRKPSLASLSGYTPR	302
OY	332 K 332		
Db	303 K 303		

RESULT 4
TF3A_XENBO STANDARD: PRT; 339 AA.

ID	TF3A_XENBO	STANDARD:	PRT;	339 AA.
AC	P17842;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-NOV-1990 (Rel. 15, Last annotation update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Transcription factor IITA (Factor A) (TFIITA).			
OC	Xenopus borealis (Kenyan clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Ampibibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCHI_TaxID=8354;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RC	MEDLINE=90245658; PubMed=2110661;			
RA	Gaskin C.J., Hanas J.S.;			
RT	"Sequence variation in transcription factor IITA."			
RL	Nucleic Acids Res. 18:2117-2123(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=90235278; PubMed=2331751;			
RA	Joho K.E., Darby M.K., Crawford E.T., Brown D.D.;			
RT	"A finger protein structurally similar to TFIITA that binds			
RL	Cell 61:293-300(1990).			
CC	-1- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR			

[illegible]

```

RESULT 5
TF3A_BUFA  STANDARD: PRT: 339 AA.
ID TF3A_BUFA
AC P34694;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 01-FEB-1995 (rel. 31, Last annotation update)
DE Transcription factor I11A (Factor A) (TFI11A).
OS Bufo americanus (American toad).
OC Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae;
OC Bufo.
OX NCBI_TaxID=8389;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=93013035; PubMed=1398134;
RA Gaskins C.J., Smith J.F., Ogilvie M.K., Hanas J.S.;
RT *Comparison of the sequence and structure of transcription factor
RL I11A from Bufo americanus and Rana pipiens.*;
CC CC
CC -I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA'S.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58366; CAA14259.1; -.
DR PIR: JCI442; JCI442.
DR HSSP: P03001; 1TF6.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 9.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_2; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW RNA-binding; Repeat; Nuclear protein.
CC CC
CC DOMAIN 12 274 ZINC FINGERS.
CC FT ZN_FING 12 36 C2H2-TYPE.
CC FT ZN_FING 42 66 C2H2-TYPE.
CC FT ZN_FING 72 97 C2H2-TYPE.
CC FT ZN_FING 104 128 C2H2-TYPE.
CC FT ZN_FING 134 158 C2H2-TYPE.
CC FT ZN_FING 161 187 C2H2-TYPE.
CC FT ZN_FING 190 212 C2H2-TYPE.
CC FT ZN_FING 219 244 C2H2-TYPE.
CC FT ZN_FING 250 274 C2H2-TYPE.
CC FT ZN_FING 250 274 C2H2-TYPE.
CC SEQUENCE 339 AA; 39500 MW; 24CE3252696183C8 CRC64;
Query Match 50.9%; Score 1022; DB 1; Length 339;
Best Local Similarity 56.1%; Pred. No. 1.8e-71;
Matches 189; Conservative 41; Mismatches 97; Indels 10; Gaps 4;
QY 34 PALPRFICSPDCSANYSKAMKLAHLCKHTGERPVCDDYBSCGAFIRDYHLSHILIT 93
DB 6 PYYKRFICSPDCNATYKKNRKLQAHLCKHGERPFCPTYESCEGFGYTLHLNHHVLS 65
QY 94 HNGEKPFVCAATGCDKFTKSNLKKHFKRKHENOOKOYICSEDDCKTKFKKQOLKIHQ 153
DB 66 HNGEKCKCKETENCNLAFTTASNMRLHFKRAHSSPAOVYCFADCGOQFRKHNLKIHQ 125
QY 154 CONTNEPFCCTOEGCGKHFASPSKLKRAKAEHGYVCKG--CSFVAKTWTELLKHVE 211

```

```

DB 126 YHTNQPFKCSHEGCDKCYASPSRLKREKTHAGYPCRKDSTCPVGVKTSWYMMNAE 185
QY 212 THKEETLCEVCRKTKFRKDYLKQHKMTNAPDERDVCRCPEGGRITYTVFNLOSHLSFH 271
DB 186 LH-SEVTCISICNTEFRKRSFKLKEHKRIHREERIVYCRPRENCDRITYTFNLKSHILTFH 244
QY 272 EESRPVPCCHAGCGKTFAMKOSLTRHAYVHDPKMKMLKYKRSREKSLASLSGYIP- 330
DB 245 ENLRPFVCEHEGCGKTFAMKOSLTRHFNTHDPEKKM---VAPRPVRSLSRLSGYKPK 301
QY 331 ---PKRQCGSLSCNGESPNCVEDKMLSTVAVLTL 364
DB 302 KSKKKKPSQTPAMESQEQPDASKADPLVLENLTL 338
RESULT 6
TF3A_ICTPU  STANDARD: PRT: 322 AA.
ID TF3A_ICTPU
AC P79797;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUL-1998 (rel. 36, Last annotation update)
DE Transcription factor I11A (Factor A) (TFI11A).
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=98086209; PubMed=9426240;
RA Ogilvie M.K., Hanas J.S.;
RT *Molecular biology of vertebrate transcription factor I11A: cloning
RL and characterization of TFI11A from channel catfish oocytes.*;
CC CC
CC -I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA'S (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z68499; CAA92808.1; -.
DR HSSP: P03001; 1TF6.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 9.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_2; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW RNA-binding; Repeat; Nuclear protein.
CC CC
CC DOMAIN 12 273 ZINC FINGERS.
CC FT ZN_FING 12 36 C2H2-TYPE.
CC FT ZN_FING 42 64 C2H2-TYPE.
CC FT ZN_FING 70 95 C2H2-TYPE.
CC FT ZN_FING 102 126 C2H2-TYPE.
CC FT ZN_FING 132 156 C2H2-TYPE.
CC FT ZN_FING 159 184 C2H2-TYPE.
CC FT ZN_FING 188 211 C2H2-TYPE.
CC FT ZN_FING 218 243 C2H2-TYPE.
CC FT ZN_FING 249 273 C2H2-TYPE.
CC SEQUENCE 322 AA; 37366 MW; EC30CE6E85F8456 CRC64;

```

```

Query Match          40.6%  Score 816; DB 1; Length 322;
Best Local Similarity 46.3%; Pred. No. 1.1e-55;
Matches 146; Conservative 55; Mismatches 92; Indels 22; Gaps 5;

OY 38 RRFICSPDCSANYSKAKMLDAHLCKHNGERPFVCDYEGCGAFLRDYHLSHILTHGE 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 10 KMFVCSFLNCKKSFSAKMLLEAHYCKHGLRPFACD--RCDTFTCRCLDTHNLHSHSGK 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 98 KPFVCAATGCDKQFNTKSLKKHFERKHEKNOQOYICSPEDCKTEKKKHOQICOCOT 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 KRYOGLDEGCSSTFSTACLKNHVERVHOKHKNHVCYEGCAKEFRKKQLRSKCELM 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 158 NEPLKCTQEGCGGKHPASPSKLKRNAAKHEGYVC--QKGSFVAKTWTLLKHVRETHKEE 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 128 NOLPEECQEGCGGKKYTKSLQKHEKVDYVPCAEBCDFOGRMWTXYQAHKKAHREA 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 217 ILCEVCRKTFRRKKOYLKOH--MKTNAPERDYCGRPREGCGRTTYTFVFNLSHLSHEER 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 LQCDSCAKYFNHAKMLKRLKFLVHLGVRVRFKCTEGCGKTYTTHFNLSHLSHEGIR 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 276 PVECEHAGCGKTFAMKQSLTRHAYVNDPDKKMKLKVKKSRKSLASHLSGYIPPKRK- 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 SPTCEHDCGGAFFAMEGSLKRHAYVNDPOKKLQKTKRGRKK-----LEPKTV 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 335 -----QGQGLSL 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 299 SDDSELPAQLHGLSL 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
P43_XENLA          STANDARD;          PRT;          365 AA.
ID P25456;
AC 01-MAY-1992 (rel. 22, Last sequence update)
DT 01-MAY-1992 (rel. 22, Last sequence update)
BT 01-FEB-1994 (rel. 28, Last annotation update)
DE P43 5S RNA binding protein (42S P43) (Thesaurin B).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90235278; PubMed=2331751;
RA Joho K.E., Darby M.K., Crawford E.T., Brown D.D.;
RT "A finger protein structurally similar to TFIIT1A that binds
   exclusively to 5S RNA in Xenopus."
RT Cell 61:293-300(1990).
RL -1- FUNCTION: P43 IS A 5S RNA BINDING PROTEIN WHICH IS A MAJOR
   CONSTITUENT OF OOCYTES AND COMPRISES PART OF A 42S
   RIBONUCLEOPROTEIN STORAGE PARTICLE.
CC -1- SUBUNIT: THE 42S RNP PARTICLE COMPRISES FOUR SUBUNITS EACH OF
   WHICH CONTAINS ONE MOLECULE OF 5S RNA, THREE MOLECULES OF TRNA,
   TWO MOLECULES OF P50 (EPI-ALPHA) AND ONE MOLECULE OF THE 5S RNA
   BINDING PROTEIN 43.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation-
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).
CC -----
CC EMBL, M32473; AAA49714.1; .
CC PIR, C34895; C34895.
CC HSSP, P03001; 1TF3.
CC InterPro: IPR000822; znf_C2H2.
CC Pfam: PF00096; zfc_C2H2; 9.
CC PRINTS: PR00048; ZINCFINGER.
CC SMART: SM00355; znf_C2H2; 9.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.

```

```

DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
KW Zinc-finger; Metal-binding; RNA-binding; Repeat.
FT DOMAIN 15 275 ZINC_FINGERS.
FT ZN_FING 15 39 C2H2-TYPE.
FT ZN_FING 45 69 C2H2-TYPE.
FT ZN_FING 75 100 C2H2-TYPE.
FT ZN_FING 106 130 C2H2-TYPE.
FT ZN_FING 136 160 C2H2-TYPE.
FT ZN_FING 163 187 C2H2-TYPE.
FT ZN_FING 191 213 C2H2-TYPE.
FT ZN_FING 220 245 C2H2-TYPE.
FT ZN_FING 251 275 C2H2-TYPE.
SQ SEQUENCE 365 AA; 41694 MW; 2B5BE6D7192995 CRC64;

Query Match          27.7%  Score 556.5; DB 1; Length 365;
Best Local Similarity 36.1%; Pred. No. 9.5e-36;
Matches 106; Conservative 47; Mismatches 122; Indels 3; Gaps 3;

OY 42 CSPDCSANYSKAKMLDAHLCKHNGERPFVCDYEGCGAFLRDYHLSHILTHGEKPFV 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 CPAAGCKAFYRREKGLDQHMACHSGQKPMKCGIKDQKVFARKQDLKHVKRHLAKLKL 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 102 CAATGCDKQFNTKSLKKHFERKHEKNOQOYICSPEDCKTKTKKQDLKIHOCOTNPEL 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 77 CPTAGCKMTFTSTKSLSRKRLYKH--GEAVPLKCFYPGCKRSFRKRLRLHLSVSHNEPL 135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 162 FKCTQEGCGGKHPASPSKLKRNAAKHEGYVC--QKGSFVAKTWTLLKHVRETHKEE 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 SVCDYVPGCSKSSVAKLVAHOKRHRGTCYEGCOTYSPTFALQTHVKR--HPIELOCA 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 221 VCRKTFRRKDYLOMKTNAPERDYCGRPREGCGRTTYTFVFNLSHLSHEESRPVCE 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 195 ACKKPFKKSALRRKATKATKAKKRLDLPGRDCCDITFSSVFLTNHVRKHLCLDTNCP 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 281 HAGCGKTFAMKQSLTRHAYVNDPDKKMKLKVKKSRK 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 255 HSGCTRSFAMRESLLRHVLVHDPERRKILKLFVRGSPSK 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
P43_XENBO          STANDARD;          PRT;          365 AA.
ID P25066;
AC 01-MAY-1992 (rel. 22, Last sequence update)
DT 01-MAY-1992 (rel. 22, Last sequence update)
BT 01-FEB-1994 (rel. 28, Last annotation update)
DE P43 5S RNA binding protein (42S P43) (Thesaurin B).
OS Xenopus borealis (Kenyan clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83354;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90235278; PubMed=2331751;
RA Joho K.E., Darby M.K., Crawford E.T., Brown D.D.;
RT "A finger protein structurally similar to TFIIT1A that binds
   exclusively to 5S RNA in Xenopus."
RT Cell 61:293-300(1990).
RL -1- FUNCTION: P43 IS A 5S RNA BINDING PROTEIN WHICH IS A MAJOR
   CONSTITUENT OF OOCYTES AND COMPRISES PART OF A 42S
   RIBONUCLEOPROTEIN STORAGE PARTICLE.
CC -1- SUBUNIT: THE 42S RNP PARTICLE COMPRISES FOUR SUBUNITS EACH OF
   WHICH CONTAINS ONE MOLECULE OF 5S RNA, THREE MOLECULES OF TRNA,
   TWO MOLECULES OF P50 (EPI-ALPHA) AND ONE MOLECULE OF THE 5S RNA
   BINDING PROTEIN 43.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation-
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).
CC -----

```


or send an email to license@isb-sib.ch.

CC EMBL: M32471; AAA49712.1; -
DR PIR: A34895; A34895.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 9.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; Znf_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
KW Zinc-finger; Metal-binding; RNA-binding; Repeat.
FT DOMAIN 15 275 ZINC_FINGERS.
FT ZN_FING 15 39 C2H2-TYPE.
FT ZN_FING 45 69 C2H2-TYPE.
FT ZN_FING 75 100 C2H2-TYPE.
FT ZN_FING 106 130 C2H2-TYPE.
FT ZN_FING 136 160 C2H2-TYPE.
FT ZN_FING 163 187 C2H2-TYPE.
FT ZN_FING 191 213 C2H2-TYPE.
FT ZN_FING 220 245 C2H2-TYPE.
FT ZN_FING 251 275 C2H2-TYPE.
SQ SEQUENCE 365 AA: 41726 MW; D7E8C76E80B42EF2 CRC64;

Query Match 27.6%; Score 553.5; DB 1; Length 365;
Best Local Similarity 37.4%; Pred. No. 1,6e-35;
Matches 108; Conservative 48; Mismatches 130; Indels 3; Gaps 3;

QY 31 PRPRALPRRFICSPDCSANYSKAMKLDALHCKHTGEPFVCDYEGCGKAFIRDYHLSKH 90
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 6 PTPSPSGVFRCPACGCKAVYRKEGKLRDMMAGHSQKLMKCGKKCKGKMFARKRQIOKH 65
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 91 ILTHNGEKPPVCAATGCDQKFNTRKSNLKKHFERNHNOQKOYICSPEDCKKTFKKHQOLK 150
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 66 MKRHLLTKHSCPTACKMTFFSTKKSLSRHKLKH -GDVAPLKCSVPGCKRSRRKRALR 124
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 151 THOQHTNPRCLKTQEGCGKHFASPSKLRNAKAHEGYVCO -KCGSFVAKTWTTELKHV 209
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 125 IHVSEHSNPLSYCDVPGCGKSTSAKLAHNRHGRGYRCSYEDQTVSPWTALQTHN 184
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 210 REHKEILLCEVCRKTFKKRDYLRKQMKTHAPRDYCRCPRECCGTYTTFVNLQSHIIS 269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 185 KK-HPELEQCAACKPKFKKASALRRKATNAKNPLQLPCPRDCCDITFTVFNLTNHLRK 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 270 FHEESRPFCEHAGCGKTFAMKOSLTRNAHVHDPDKKKKLLKVKKSRK 318
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 244 VHLCTGHTHCPSNCTRSFAMRESLYRHLVYHDPREKKKIKLTKGRPRSK 292
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 9
ZXDB_HUMAN STANDARD; PRT; 803 AA.
AC P98169; OQUB3;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger X-linked protein ZXDB.
GN ZXDB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 307-709 FROM N.A.
RC TISSUE=Brain;
MEDLINE=94093547; PubMed=8268913;
RA Greig G.M., Sharp C.B., Carrel L., Willard H.F.;
RT "Duplicated zinc finger protein genes on the proximal short arm of
RT the human X chromosome: Isolation, characterization and
RT X-inactivation studies.";

Hum. Mol. Genet. 2:1611-1618(1993).
CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1 TISSUE SPECIFICITY: UBIQUITOUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AL031115; CAB36858.1; -
DR EMBL: Z99130; CAB16205.1; -
DR EMBL: L14788; AAC37522.1; -
DR HSPF; P03001; 1T93.
DR Genew; HGNC:13199; ZXDB.
DR MIM; 300236; -
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 10.
DR SMART; SM00355; Znf_C2H2; 10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 15 20 POLY-GLY.
FT DOMAIN 83 89 POLY-GLY.
FT ZN_FING 271 572 ZINC_FINGERS.
FT ZN_FING 271 295 C2H2-TYPE.
FT ZN_FING 304 328 C2H2-TYPE.
FT ZN_FING 334 358 C2H2-TYPE.
FT ZN_FING 364 386 C2H2-TYPE.
FT ZN_FING 393 417 C2H2-TYPE.
FT ZN_FING 424 448 C2H2-TYPE.
FT ZN_FING 454 478 C2H2-TYPE.
FT ZN_FING 484 508 C2H2-TYPE.
FT ZN_FING 514 538 C2H2-TYPE.
FT ZN_FING 547 572 C2H2-TYPE.
SQ SEQUENCE 803 AA: 84792 MW; FDS46D326FF5A5D4 CRC64;

Query Match 20.8%; Score 417.5; DB 1; Length 803;
Best Local Similarity 29.1%; Pred. No. 9,8e-25;
Matches 111; Conservative 47; Mismatches 131; Indels 93; Gaps 14;

QY 18 DATAIGESSAPP-----PRPALPR-----FICSPDCSANYSKAMKLD 58
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 230 DLLAPRAPAPAPAPAPAPAPAPALGPRGLSGGPGVVLICPAPAGCGQTFARKHQLK 289
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 59 AHLCKHT---GERPFVCDYEGCGKAFIRDYHLSRHLITHTGKPFVCAATGCDQKFNTRK 115
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 290 VHLTHTSSQGGPRFCPLGGCGWTFSTYKLRHLQSHDKLRFPGCPAEGGCKSFTTYV 349
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 116 NLKKHF-----ERKHNQKOYICSPEDCKKTFKKHQO 148
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 350 NLKANHKGHOENSFCEVCEESFPTQAKLSAHOBFHFERERYQCAFSGCKHTFTVSA 409
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 149 LKTHOQHTNE-PLFKCTQEGCGKHFASPSKLRNAKAHEG--YYCQ-KGCSFVAKTWT 203
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 410 LFSHNAHREDELFSCSPGCKQYDKACRLKTLHRSHTGERPFLCDFDGCWMTTSHS 469
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 204 ELKHVRETHKEILLCEVCRKTFKKRDYLRKQMKTHAPRDYCRCPRECCGTYTTFVNL 263
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 470 KILRH-----KRK-----HDDDRRF-MCPVEGCGKSFTAEHL 501
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 264 QSHILSFHEESRPFVCEHAGCGKTFAMKOSLTRNAHVH--DPDKKKMKLV-----KKSRE 317
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 502 KGHSLIT-HLGTKPFVCPVAGCCARFSARSLYTHSKKHLQVDVTKWRSRCPISSCNKLFTS 560
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 318 KRSLASHSLGYIPPKRKQOGL 339
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 561 KHSMTKTHM-----VKRKAVGDL 578
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 10

DR SMART: SM00355; ZNF_C2H2; 12.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 12.
KM Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT NON_TER 1 1
FT ZN_FING 6 28 C2H2-TYPE.
FT ZN_FING 34 56 C2H2-TYPE.
FT ZN_FING 62 94 C2H2-TYPE.
FT ZN_FING 90 112 C2H2-TYPE.
FT ZN_FING 118 140 C2H2-TYPE.
FT ZN_FING 146 168 C2H2-TYPE.
FT ZN_FING 174 196 C2H2-TYPE.
FT ZN_FING 202 224 C2H2-TYPE.
FT ZN_FING 230 252 C2H2-TYPE.
FT ZN_FING 258 280 C2H2-TYPE.
FT ZN_FING 286 308 C2H2-TYPE.
FT NON_TER 336 336
SQ SEQUENCE 336 AA; 37554 MW; 815B667602330B46 CRC64;

Query Match 20.4%; Score 409; DB 1; Length 336;
Best Local Similarity 34.0%; Pred. No. 1.8e-24;
Matches 108; Conservative 38; Mismatches 122; Indels 50; Gaps 15;

QY 40 FICSPDSCANSKAWKLDANLCKHTGERPVCDYEGCGKAFIRDYHLSRHILTHTEGKP 99
DB 34 FICT--ECGKGFSOKGILOTNMTHTGKERPTC--TECGKNFAQIYTLRLHILHTGKERP 89
QY 100 FYCANTGCDOKNTKSNLKKHFERKHENOKOYICSEFDECKTKTKKHOOLKTHOCQHTNE 159
DB 90 FSC--TECGKHFAHKGHLVSHM--KTHTEGKERPTC--ECGKHFAQGHVSHMKTHTGE 143
QY 160 PLEKTOEGCGKHFAPSPKLRHAKANEH--YVCOK--GCSFYAKTWTELLKHYR--ETNK 214
DB 144 KPTCTE--CGKNFAQKMTLCHLKHTEGKERPTCTEGCDKFAKK--NNILRLKLTHTGE 199
QY 215 EELICEVCKTKRKDYLOKHMKTNAERDVSCPRGCGKRTYTVFNLSHLSFEHES 274
DB 200 KPTCTEGCGKAFITLKGSLVGHMKHTGKERPT--SCIQ--CGKNFTOKNSLICH--LTNHTGE 255
QY 275 RPPVCEHAG-----CGKTFAMKOSLTRAVVNDPRKKM 308
DB 256 KPTCTEGCGKGFALGNLVLHTKHTGKERPTSCGCKNFAQKNSLRLKLTHTREKPT 315
QY 309 KLVK--KKSREKRSLSHL 325
DB 316 YSECGKKYSQIYNLASHM 333

RESULT 12
XZDA_HUMAN STANDARD; PRT; 799 AA.
AC P98168: OGU77;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger X-linked protein ZXDA.
GN ZXDA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 303-759 FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=94093547; PubMed=8268913;
RA Greig G.M., Sharp C.B., Carrel L., Willard H.F.:
RT "Duplicated zinc finger protein genes on the proximal short arm of
RT the human X chromosome: Isolation, characterization and
RT X-inactivation studies.";

RL Hum. Mol. Genet. 2:1611-1618(1993).
CC -I- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -I- TISSUE SPECIFICITY: UBINOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL034396; CAB46717.1; -
CC EMBL: L14787; AAC37521.1; -
CC HSSP: P03001; TRF3.
CC Genew; HGNC:13198; ZXDA.
CC MIM: 300235; -
DR InterPro: IPR000822; ZNF_C2H2.
DR Pfam: PF00096; zf-C2H2; 10.
DR SMART: SM00355; ZNF_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 9.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein.
KW DOMAIN 15 20 POLY-GLY.
FT ZN_FING 267 568 ZINC_FINGERS.
FT ZN_FING 267 291 C2H2-TYPE.
FT ZN_FING 300 324 C2H2-TYPE.
FT ZN_FING 330 354 C2H2-TYPE.
FT ZN_FING 360 382 C2H2-TYPE.
FT ZN_FING 389 413 C2H2-TYPE.
FT ZN_FING 420 444 C2H2-TYPE.
FT ZN_FING 450 474 C2H2-TYPE.
FT ZN_FING 480 504 C2H2-TYPE.
FT ZN_FING 510 534 C2H2-TYPE.
FT ZN_FING 543 568 C2H2-TYPE.
FT CONFLICT 753 759 AEMSVHP -> SGVERTS (IN REF. 2).
SQ SEQUENCE 799 AA; 84770 MW; 112BF6C6D0CA670 CRC64;

Query Match 20.4%; Score 409; DB 1; Length 799;
Best Local Similarity 29.2%; Pred. No. 4.4e-24;
Matches 115; Conservative 48; Mismatches 137; Indels 94; Gaps 15;

QY 4 PAVVAESVSLTADAFIAGESSAPTPRPAL-----PRFICSPD----- 46
DB 217 PGDCPELRSDLLAE--PAEPAPAPQDEAEGLAALGPRLOLSSGCVLYLCEPAL 273
QY 47 CSANYSKAWKLDANLCKHT--GERPVCDYEGCGKAFIRDYHLSRHILTHTEGKPPVCA 103
DB 274 CGOTFAKKHQLKMHLLTHSSOGORPFCPLGCGGWTFTTSYKLKRHLSDKLRPFCGP 333
QY 104 ATGCCDCKFTKSLKKNF-----ERKHENOOKOYICSF 136
DB 334 AEGCGKSFYTVNVLKAMKMGHEOENSFKCEVCESESPFOAKLGAHQRSHPEPRPQCAF 393
QY 137 EDCKTKFKKHQOLKIQCQHTNE--PLEKTOEGCGKHFAPSPKLRHAKANEH--YVCO 192
DB 394 SGCKKTFITVSALFSSNRAHFRQDELFCSPGCSQYDKACYLKHLNLSHTGERFLCD 453
QY 193 -KGCSTVAATWTELLKHNVBETHKEELICEVCRTERKRDYLOKHMKTNAERDVSCRPRE 251
DB 454 FDCGWNFTSMSKLRLH-----HDDDRRF--MCPVE 485
QY 252 GCGRTTYTVFNLSHLSFEHESRPVCEHAGCGKTFAMKOSLTRAVVND--DPRKRRK 309
DB 486 GCGKSFTRAHNLKGHSLT--HLGKRPVCPVAGCCARFSAASSLYHSKRLADVDVTKSR 544
QY 310 LKV----KKSREKRSLSHLSGYTPPKRKGQGL 339
DB 545 CPISSCNKLFTSKHSMKTHM-----VKRHVYGQDL 574

RESULT 13
Z234_HUMAN

```

ID 2234_HUMAN STANDARD; PRT; 698 AA.
AC Q14588; Q9NS45; Q9NTT7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 234 (zinc finger protein HZF4) (Fragment).
GN ZNF234 OR ZNF269.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-73 FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Stubbs L.;
RT "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 6-698 FROM N.A.
RA MEDLINE=95169271; PubMed=7865130;
RX Abtink M., Aveskog M., Hellman L.;
RA "Isolation of cDNA clones for 42 different Kruppel-related zinc finger
RT proteins expressed in the human monoblast cell line U-937."
RL DNA Cell Biol. 14:125-136(1995).
RN [3]
RP SEQUENCE OF 7-698 FROM N.A.
RA Kodoyanni V., Ge Y., Severin J., Krummel G.K., Grabbe L.,
RA Kvalstad E., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT "Sequence analysis of a 1kb region in 19q13.2 containing a zinc finger
RT gene cluster."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF228417; AAF34785.1; -
DR EMBL: X78927; CAN55527.1; -
DR EMBL: AC074331; AAF8104.1; -
DR HSSP: P08047; 1SP2.
DR Genew: HGNC:13027; ZNF234.
DR MIM: 604750; -
DR InterPro: IPR00109; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 19.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINCfinger.
DR ProDom: PD000003; Znf_C2H2; 5.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 19.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 19.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 8 90 KRAB.
FT DOMAIN 91 175 KRAB.
FT DOMAIN 176 198 ZINC_FINGERS.
FT ZN_FING 176 198 C2H2-TYPE.
FT ZN_FING 204 226 C2H2-TYPE.
FT ZN_FING 232 254 C2H2-TYPE.
FT ZN_FING 260 282 C2H2-TYPE.
FT ZN_FING 288 310 C2H2-TYPE.

```

```

FT ZN_FING 316 338 C2H2-TYPE.
FT ZN_FING 344 366 C2H2-TYPE.
FT ZN_FING 372 394 C2H2-TYPE.
FT ZN_FING 400 422 C2H2-TYPE.
FT ZN_FING 428 450 C2H2-TYPE.
FT ZN_FING 456 478 C2H2-TYPE.
FT ZN_FING 484 506 C2H2-TYPE.
FT ZN_FING 512 534 C2H2-TYPE.
FT ZN_FING 540 562 C2H2-TYPE.
FT ZN_FING 568 590 C2H2-TYPE.
FT ZN_FING 596 618 C2H2-TYPE.
FT ZN_FING 624 646 C2H2-TYPE.
FT ZN_FING 652 674 C2H2-TYPE.
FT CONFLICT 494 494 N -> S (IN REF. 3).
FT CONFLICT 600 600 A -> E (IN REF. 3).
FT NON_TER 698 698
SQ SEQUENCE 698 AA; 80271 MW; 936D390D02E6746B CnC64;

Query Match 20.2%; Score 406; DB 1; Length 698;
Best Local Similarity 30.9%; Pred. No. 6-5e-24;
Matches 105; Conservative 42; Mismatches 119; Indels 74; Gaps 14;

OY 46 DCSANYSKAMKIDALHCKHTGERPVCDEGCGAFIRDYHLSRIHLHTGKRPVCAAT 105
DB 236 ECGKGFSTRSTLVHCKHLSGKPYNC--EECGAFIHASHIOEQIRIHTGKRPKCDT- 292
OY 106 GCDOKFNTRSNLKKHPERKHENDQKQYICSPEDCKKTKKHQOLKHOCQHTNEPLFCT 165
DB 293 -CGKPFRRRSALNNHC--MVHTGKPYKC--EDCGKCFCSNLHIOHVRHTGKPYKC- 346
OY 166 OEGCKHFPASPSKLRHAKAHNG---YVQC---KG-----CSFV 198
DB 347 -EECKGCFIOPQFOAHRIHTGKPYCKGKCFIYSSFOAHQGVHTGKPYKNC 405
OY 199 ATWTBELKH-----VRETHKEELCEVCKTKRKDYLKQHKHTAPREDVRCRREGG 254
DB 406 GKSFRMKIHYQVHLVHTGKPYKCEVCGKAFRQSSYLKIHKAHSVQPF-KC--EECG 462
OY 255 RYTVVFNLSHILSFHEESRPVCEHAG-----CGKTF 288
DB 463 QGFNOSRQIHDL-IHTGKPYKCECGKGFRRRADLKHCRHTGKPYNCEGKYF 521
OY 289 AKKSLTRHVVHDPDKKMKLKKSKRSKSLASHSGY 328
DB 522 SQASHLTHQHVHSGEK--PRKCECGKSFSAHLQAH 558

RESULT 14
ZNF41_HUMAN STANDARD; PRT; 821 AA.
AC P51814; Q9UMC4; Q9GLE8; Q9UMV6; Q9UMV7; Q9UMV8; Q9UMV9;
AC Q9UMW0; Q9UMW1;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 41.
GN ZNF41.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99380103; PubMed=10449920;
RA Rosati M., Franze A., Matrazzo M.R., Grimaldi G.;
RT "Coding region Intron/exon organization, alternative splicing and
RT x-chromosome inactivation of the KRAB/FPB-domain-containing human zinc
RL finger gene ZNF41."
RN [2]
RP CytoGenet. Cell Genet. 85:291-296(1999).
RP SEQUENCE OF 280-821 FROM N.A.
RX MEDLINE=9124317; PubMed=2037297;
RA Franze A., Archidiacono N., Rocchi M., Marino M., Grimaldi G.;

```

FT		*Isolation and expression analysis of a human zinc finger gene
Rt		(ZNF41) located on the short arm of the X chromosome.";
RL	Genomics 9:728-736(1991).	
RN	[3]	
RP	SEQUENCE OF 142-821 FROM N.A.	
RA	Brown A.:	
RL	Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.	
CC	- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.	
CC	- SUBCELLULAR LOCATION: Nuclear (potential).	
CC	- ALTERNATIVE PRODUCTS: At least 8 isoforms; 1 (shown here), 2, 3,	
CC	4, 5, 6, 7 and 8; are produced by alternative splicing.	
CC	- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-	
CC	FINGER PROTEINS.	
CC	- SIMILARITY: CONTAINS 1 KRAB DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; X60155; CAB51740.1; -	
DR	EMBL; AJ010017; CAB53035.1; -	
DR	EMBL; AJ010018; CAB53036.1; -	
DR	EMBL; AJ010019; CAB53037.1; -	
DR	EMBL; AJ010020; CAB53038.1; -	
DR	EMBL; AJ010021; CAB53039.1; -	
DR	EMBL; AJ010022; CAB53040.1; -	
DR	EMBL; AJ010023; CAB53041.1; -	
DR	EMBL; M92443; AAA61312.1; -	
DR	EMBL; AL590223; CAC8162.1; -	
DR	HSSP; P25490; 1ZNM.	
DR	Newseq; HGNC:13107; ZNF41.	
DR	MIM; 314995; -	
DR	InterPro; IPR001909; KRAB.	
DR	InterPro; IPR008822; ZnF_C2H2.	
DR	Pfam; PF00096; zf-C2H2; 51.	
DR	Pfam; PF01352; KRAB; 5.	
DR	Prodrom; PD000003; ZnF_C2H2; 14.	
DR	SMART; SM00349; KRAB; 1.	
DR	SMART; SM00355; ZnF_C2H2; 17.	
DR	PROSITE; PS50805; KRAB; 1.	
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.	
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 18.	
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding; Nuclear protein; Repeat; Alternative splicing.	
FT	DOMAIN 69 140 KRAB.	
FT	DOMAIN 313 811 ZINC FINGERS.	
FT	ZN_FING 313 335 C2H2-TYPE.	
FT	ZN_FING 341 364 C2H2-TYPE (DEGENERATE).	
FT	ZN_FING 369 391 C2H2-TYPE.	
FT	ZN_FING 397 419 C2H2-TYPE.	
FT	ZN_FING 425 447 C2H2-TYPE.	
FT	ZN_FING 453 475 C2H2-TYPE.	
FT	ZN_FING 481 503 C2H2-TYPE.	
FT	ZN_FING 509 531 C2H2-TYPE.	
FT	ZN_FING 537 559 C2H2-TYPE.	
FT	ZN_FING 565 587 C2H2-TYPE.	
FT	ZN_FING 593 615 C2H2-TYPE.	
FT	ZN_FING 621 643 C2H2-TYPE.	
FT	ZN_FING 649 671 C2H2-TYPE.	
FT	ZN_FING 677 699 C2H2-TYPE.	
FT	ZN_FING 705 727 C2H2-TYPE.	
FT	ZN_FING 733 755 C2H2-TYPE.	
FT	ZN_FING 761 783 C2H2-TYPE.	
FT	ZN_FING 789 811 C2H2-TYPE.	
FT	VARSPLIC 1 60 MAANDSPPMVALAAEGRGSCVEVRREPTARIHSVKRY PDLSGPKGRSSAHAAIN --> MGTLPHGTRPMLORVYAA HV (IN ISOFORM 7 AND ISOFORM 2). MAANDSPPMVALAAEGRGSCVEVRREPTARIHSVKRY FDLSGPCGRGS --> MGTLPHGPFWLORDVAHHV (IN	
FT	VARSPLIC 1 52	

```
CC FT VARSPLIC 25 66 MISSING (IN ISOFORM 3) .
CC FT VARSPLIC 53 60 MISSING (IN ISOFORM 5) .
CC FT VARSPLIC 1 128 MISSING (IN ISOFORM 4) .
CC FT VARSPLIC 141 176 MISSING (IN ISOFORM 7 AND ISOFORM 8) .
CC FT CONFLICT 111 111 Q -> R (IN REF. 1 : CAB53039) .
CC SO SEQUENCE 821 AA; 93728 MW; F4AB7B0BC9A8AF13 CACG64 ;
CC
CC Query Match Best Local Similarity 20.1% Score 404; DB 1: Length 821;
CC Matches 106; Conservative 46; Mismatches 121; Indels 52; Gaps 16;
CC
CC Oy 42 CSFPDCSANYSKAMKLDLADHLCKHTGEERPEFYDDECGCAFLRDYLSNHIILTHNGEKPV 101
CC Db 511 CS--DCGSFFTKSSLJAHORIHTEKEPYIC-TECGAVFHRTNLTTNOHXTKHCEPKYM 566
CC Oy 102 CAATGCDDCFMPTNSLYLKHFERNHNQOKOYCSEFEDCKPFKKHQQLIKHOCTNPPL 161
CC Db 567 CAE-CGAFAFDOSULIHH-OXITHEGPXYCN-GCGKAFTMSKRLIKHOXSIGERRH 620
CC Oy 162 FKCTGGCGRHKHPASPSKLRRNAKAHEG---YYOCR-GCSPAKTWTELLEKNVR-ETHERE 216
CC Db 621 YEC-KDGCAFIOQSSTLVHORITHGEPRVCPCBGAKFIQG-SHFIANRHITGEXP 676
CC Oy 217 ILCEVCRTTFRKDYDKLOMKTNADER-DVCECPREGSGRTTYPFNLOSILSHESNR 275
CC Db 677 YECSOGCGKFTRKSOLRYHOKITHGEKENIC---AECCKAFTDRSNLIITH-OXITHREK 731
CC Oy 276 PVVC-----EHAG-----CGKTPFMQSILTRAHVVDPPDKMMK 309
CC Db 732 PEECDGCRGTFWKSRRLNIHOKSHGERHYECSKGCAFKAIOTATLSMOIIHTGKKPVAC 791
CC Oy 310 LKVKSREKRS-LASHLSGYIPPKR 333
CC Db 792 TBCKAKFTDRSLIRKHOKMHSGEK 816
CC
CC RESULT 15
CC YE73_HUMAN STANDARD; PRt: 574 AA.
CC AC ODP255;
CC DT 15-JUN-2002 (Rel. 41, Created)
CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Hypothetical zinc finger protein KIA1473 (Fragment).
CC GN KIA1473.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC RP SEQUENCE FROM N.A..
CC RC TISSUE=Brain;
CC RX MEDLINE=20277482; PubMed=10819331;
CC RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
CC RT "Prediction of the coding sequences of unidentified human genes. XVII.
CC RT The complete sequences of 100 new cDNA clones from brain which code
CC RT for large proteins in vitro.";
CC RL DNA Res. 7:143-150(2000).
CC CC -1 FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC CC -1 SUBCELLULAR LOCATION: Nuclear (Probable).
CC CC -1 SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC CC -1 SIMILARITY: CONTAINS 1 KBAB DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isdb-sib.ch/announce/cisb-stb.ch).
```

DR EMBL: AB040906; BAA95997.1; .
DR HSSP: P07248; 1AR0.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; ZF-C2H2; 13.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; ZnF_C2H2; 2.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; ZnF_C2H2; 13.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT NON_TER 1 1
FT DOMAIN 15 86 KRAB.
FT DOMAIN 184 542 ZINC_FINGERS.
FT ZN_FING 184 542
FT ZN_FING 184 206 C2H2-TYPE.
FT ZN_FING 212 234 C2H2-TYPE.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 268 290 C2H2-TYPE.
FT ZN_FING 296 318 C2H2-TYPE.
FT ZN_FING 324 346 C2H2-TYPE.
FT ZN_FING 352 374 C2H2-TYPE.
FT ZN_FING 380 402 C2H2-TYPE.
FT ZN_FING 408 430 C2H2-TYPE.
FT ZN_FING 436 458 C2H2-TYPE.
FT ZN_FING 464 486 C2H2-TYPE.
FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).
FT ZN_FING 520 542 C2H2-TYPE.
SQ SEQUENCE 574 AA; 65951 MM; 48255F1176CASA80 CRC64;

Query Match 20.1%; Score 403; DB 1; Length 574;
Best Local Similarity 31.0%; Pred. No. 8.9e-24;
Matches 106; Conservative 47; Mismatches 121; Indels 68; Gaps 17;

OY 46 DCSANYSKAMKIDALHCKHTGERPVCDEGCGKAFIRDYHLSRHILFTTGEKPFVCAAT 105
DB 216 ECGKAVNETSNLSTHKRIHTGKKPKYC--EECGKAFNRLSHLTTHKRIHTGKKPKCEE- 272
OY 106 GCDQKFNKSNLKKHFERKHENO-----OKOYICSEFD 138
DB 273 -CGKAFNOSANLTTN-KRIHTGEKPYKCECGRAFSSQSTLTANHRIHAGEKPYKC--EE 328
OY 139 CKKTKKKHQQLKHOCQHTNEPLFKCTOGCGCKHAFSPSKLKRHAKHNGC---YVQCKGC 195
DB 329 CGKAFSSQSTLTTHKRIHTGEKPYKC--EECGKAFSOLSHLTTHKRIHSGEKPYKCEB-C 385
OY 196 SPVAKTWTELLKHAVRETHKEIL-CEVCRKTKFRKDYLKQNHKTHAPERDVCRCPREGCG 254
DB 386 GKAFKQSSLTTHKRIHAGEKPYKCEVCGSKAFSRSHLTTHKRIHTGEKPY-KC--EECG 442
OY 255 RYTTVPFNLQSHILS---FHESRPVCEHAGCGKTFANKQSLTRHVVHDPDKKKMKLK 311
DB 443 K---AFNLSSQLTTHKRIHTGEKPYKCEE--CGKAFNQSSTLSKKHVIHTGEC---PYK 493
OY 312 VKKSREKRLASHLSGYTPPKRKQGGSLSCNGESP-NCYE 352
DB 494 YECCGKAFNQSSTLTTH-----KMIHTGEKPYKCEE 524

Search completed: February 10, 2003, 17:45:49
Job time : 26 secs

Matches 1206; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```

QY 1 GTGCGGCGCGCGCGGCGGAGGTTTCAGCAGGAGCGTGCGCGCGCGCGCGGTTCCCGGC 60
Db GTGCGGCGCGCGCGCGGCGGAGGTTTCAGCAGGAGCGTGCGCGCGCGCGCGGCGGCG -CGGTTCCCGGC 138
QY 61 ACCTGTCGCGCACGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 139 ACCTGTCGCGCACGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 138
QY 121 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 199 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258
QY 181 AGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 259 AGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 318
QY 241 CTGCTTCCTGACTGACGCGCGCAATTACAGCAAAAGCCTGGAGGCTTGACGCGACCTGTG 300
Db 319 CTGCTTCCTGACTGACGCGCGCAATTACAGCAAAAGCCTGGAGGCTTGACGCGACCTGTG 378
QY 301 CAAGCAGCGCGGAGAGGACGACATTTGTTGTGACTATGAAGGCTGTGCGAAGCGCTTCAT 360
Db 379 CAAGCAGCGCGGAGAGGACGACATTTGTTGTGACTATGAAGGCTGTGCGAAGCGCTTCAT 438
QY 361 CAAGGACTACCATCTGAGCGCGCGCATTTGACTCACACAGGAGAAAGCGTTGTTGTTG 420
Db 439 CAAGGACTACCATCTGAGCGCGCGCATTTGACTCACACAGGAGAAAGCGCTTTGTTGTTG 498
QY 421 TGCAGCCACTGGCTGTGATCAAAAATTCAACACAAAATCAAACTGGAAGAACATTTTGA 480
Db 499 TGCAGCCAAATGGCTGTGATCAAAAATTCAACACAAAATCAAACTGGAAGAACATTTTGA 558
QY 481 ACGCAAAATGAAATTAACAAAAACATATATATGCGATTTTGAAGCTGTGAAGAAC 540
Db 559 ACGCAAAATGAAATTAACAAAAACATATATATGCGATTTTGAAGCTGTGAAGAAC 618
QY 541 CTTTAAAGAAATCAGCAGGCTGAAATTCATCAGTGGCAGCATCCAGTACCTCATTT 600
Db 619 CTTTAAAGAAATCAGCAGGCTGAAATTCATCAGTGGCAGCATCCAGTACCTCATTT 678
QY 601 CAAGTGTACCGAGAGGATGTGGGAAACATTTGATCAACCGCAGCAAGCTGAAACGACA 660
Db 679 CAAGTGTACCGAGAGGATGTGGGAAACATTTGATCAACCGCAGCAAGCTGAAACGACA 738
QY 661 TGGCAAGCGCGCGAGAGGCTATGATGTCAAAAAGATGTTCTTTTGGCAAAAACATG 720
Db 739 TGGCAAGCGCGCGAGAGGCTATGATGTCAAAAAGATGTTCTTTTGGCAAAAACATG 798
QY 721 GAGGAACTTCTGAAACATGTGAGAGAACCCATAAAGAGAAATACTATGTGAAGTATG 780
Db 799 GAGGAACTTCTGAAACATGTGAGAGAACCCATAAAGAGAAATACTATGTGAAGTATG 858
QY 781 CCGGAAAAATTTAAACGCAAGATTAACCTTAAGCAACACATGAAAACTATGCCCGACA 840
Db 859 CCGGAAAAATTTAAACGCAAGATTAACCTTAAGCAACACATGAAAACTATGCCCGACA 918
QY 841 AAGGAGATGATGTGCGCTGTCAAGAGAAAGGCTGTGGAAGAACCTATATCTACTGTGTTAA 900
Db 919 AAGGAGATGATGTGCGCTGTCAAGAGAAAGGCTGTGGAAGAACCTATATCAACTGTGTTAA 978
QY 901 TGTCCAAAGGCAATTCCTCTCTCTCATGAGGAAAGCGCGCGCTTTTGTGTGTGAACATGC 960
Db 979 TGTCCAAAGGCAATTCCTCTCTCTCATGAGGAAAGCGCGCGCTTTTGTGTGTGAACATGC 1038
QY 961 TGGCTGTGCGCAAAACATTTGCAATGAACAAGTCTCATAGGCAATCTTTGTATACATGA 1020
Db 1039 TGGCTGTGCGCAAAACATTTGCAATGAACAAGTCTCATAGGCAATCTTTGTATACATGA 1098
QY 1021 TCGTGACAGAGAAATGAAGCTCAAAAGTCAAAAATTCGTGTGAAGAAC -GCAAGTTGG 1079
Db 1099 TCGTGACAGAGAAATGAAGCTCAAAAGTCAAAAATTCGTGTGAAGAACGCGAGTTGG 1158

```

```

QY 1080 CCTCTCATCTCAGTGGATATATCCCTCCAAAGAAAGGCAAGGCTATATCTTTGT 1139
Db 1159 CCTCTCATCTCAGTGGATATAT-CTCTCCAAAGAAAGGCAAGGCTATATCTTTGT 1217
QY 1140 GTCAAAAGGAGAGTCAACCCCACTGTGTGGAAGACAGATGCTCTGACAGTTGCAGTAC 1199
Db 1218 GTCAAAAGGAGAGTCAACCCCACTGTGTGGAAGACAGATGCTCTGACAGTTGCAGTAC 1277
QY 1200 TTACCCCTTGGCTAA 1213
Db 1278 TTACCCCTTGGCTAA 1291

```

RESULT 2
US-08-523-376-1
: Sequence 1, Application US/08523376
: Patent No. 5808030
: GENERAL INFORMATION:
: APPLICANT: Tsutomu, FUJIWARA
: APPLICANT: Satoshi, TAKEDA
: APPLICANT: Yoshikazu, SHIMADA
: APPLICANT: Kouichi, OKAZAKI
: TITLE OF INVENTION: hTFIIIA GENE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
: STREET: 2100 Pennsylvania Avenue, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: United States
: ZIP: 20037-3202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/523,376
: FILING DATE:
: CLASSIFICATION: 536
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 293-7060
: TELEFAX: (202) 293-7860
: TELEX: 6491103
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1269 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-523-376-1

Query Match 96.2%; Score 1167; DB 1: Length 1269;
Best Local Similarity 99.3%; Pred. No. 1.3e-302;
Matches 1203; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```

QY 1 GTGCGGCGCGCGCGGCGGAGGTTTCAGCAGGAGCGTGCGCGCGCGCGCGGTTCCCGGC 60
Db 61 GTGCGGCGCGCGCGGCGGAGGTTTCAGCAGGAGCGTGCGCGCGCGCGCGGCGGCG -CGGTTCCCGGC 119
QY 61 ACCTGTCGCGCACGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 120 ACCTGTCGCGCACGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
QY 121 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 180 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
QY 181 AGCGGCGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

```


Db 240 AGCCGGGAGAGCTCAGTCCGACCCGCCGCCGCTTCCAGAGAGTTCATCTG 299
QY 241 CTCCTTCCCTCAGTCAGAGCCGCAATTAACGAAGCCTGGAAAGCTTGACGGGCACCTGTG 300
Db 300 CTCCTTCCCTCAGTCAGAGCCGCAATTAACGAAGCCTGGAAAGCTTGACGGGCACCTGTG 359
QY 301 CAAGCAGACGGGGAGAGACATTTGTTGTGACTATGAAGGGTGTGGCAAGGCTTCAT 360
Db 360 CAAGCAGACGGGGAGAGACATTTGTTGTGACTATGAAGGGTGTGGCAAGGCTTCAT 419
QY 361 CAGGAGTACCATCTGAGCCGCCACATTCCTGACTCACACAGGAGAAAAGCCGTTGTTTG 420
Db 420 CAGGAGTACCATCTGAGCCGCCACATTCCTGACTCACACAGGAGAAAAGCCGTTGTTTG 479
QY 421 TGCAGCCACGTGGCTGATCAAAAATTCACAAACAAATCAACTTGAAGAAACATTTTGA 480
Db 480 TGCAGCCACGTGGCTGATCAAAAATTCACAAACAAATCAACTTGAAGAAACATTTTGA 539
QY 481 ACCGAACATGAAATATCAACAAACAAATATATATGAGTTTGAAGACTGTAAGAAGAC 540
Db 540 ACCGAACATGAAATATCAACAAACAAATATATATGAGTTTGAAGACTGTAAGAAGAC 599
QY 541 CTTTAAGAAACATGACAGACTGAAATTCATCACTGCCAGCATCAATGAACCTCTATT 600
Db 600 CTTTAAGAAACATGACAGATGAAATTCATCACTGCCAGCATCAATGAACCTCTATT 659
QY 601 CAATGTATCCAGAGAGATGTGGGAAACACTTTGCACTACCCAGCAAGCTGAACGCA 660
Db 660 CAATGTATCCAGAGAGATGTGGGAAACACTTTGCACTACCCAGCAAGCTGAACGCA 719
QY 661 TGCAGAGCCGACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 720
Db 720 TGCAGAGCCGACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 779
QY 721 GAGGAACTTCTGAACATGTGAGAGAAACCCATPAAGAGGAAATACTATGTGAATATG 780
Db 780 GAGGAACTTCTGAACATGTGAGAGAAACCCATPAAGAGGAAATACTATGTGAATATG 839
QY 781 CCGGAAACATTTAAGGCAAGATTAACCTTAAGCAACACATGAAACTCATGTGCCCGCA 840
Db 840 CCGGAAACATTTAAGGCAAGATTAACCTTAAGCAACACATGAAACTCATGTGCCCGCA 899
QY 841 AAGGATGTATGTGCTGTCAGAGAGAGGCTGTGAGAAACCTATACTACTGTGTTTAA 900
Db 900 AAGGATGTATGTGCTGTCAGAGAGAGGCTGTGAGAAACCTATACTACTGTGTTTAA 959
QY 901 TCTCCAAAGCATATCTCTCTCTTCATGAGGAAAGCCGCTTTTGTGTGTAACATGC 960
Db 960 TCTCCAAAGCATATCTCTCTCTTCATGAGGAAAGCCGCTTTTGTGTGTAACATGC 1019
QY 961 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGTCTCACTAGGATGCTGTGTAACATGA 1020
Db 1020 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGTCTCACTAGGATGCTGTGTAACATGA 1079
QY 1021 TCCAGCAAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTAAGAAAC-GGAGTTTGG 1079
Db 1080 TCCAGCAAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTAAGAAAGGGAGTTTGG 1139
QY 1080 CCTCTCATCTAGTGATATATTCCTCCCAAGAGAAACAAAGGCAAGCTTATCTTTGT 1139
Db 1140 CCTCTCATCTAGTGATATAT-TCTCCCAAGAGAAACAAAGGCAAGCTTATCTTTGT 1198
QY 1140 GTCAAAAGGAGAGTCAACCCAGTGTGGAAGCAAGATCTCTCGACATGTGACATG 1199
Db 1199 GTCAAAAGGAGAGTCAACCCAGTGTGGAAGCAAGATCTCTCGACATGTGACATG 1258
QY 1200 TTACCTTGGC 1210
Db 1259 TTACCTTGGC 1269

RESULT 3
US-09-492-985-11

Sequence 11, Application US/09492985
Patent No. 6376240
GENERAL INFORMATION:
APPLICANT: Song, An M.
APPLICANT: Chen, Ya-Fen
TITLE OF INVENTION: REFAT-1: A Transcription Factor That
FILE REFERENCE: SUN-113P
CURRENT APPLICATION NUMBER: US/09/492,985
EARLIER FILING DATE: 2000-01-27
EARLIER APPLICATION NUMBER: 60/117,576
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 1471
TYPE: DNA
ORGANISM: mouse
US-09-492-985-11

Query Match
Best Local Similarity 48.9%; Pred. No. 2e-07;
Matches 170; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 67 CTCGGCAGCTGCGACGCCGCTGGCCCTGGGCTTGAGAGGCGCCGCCCTGATCCGCC 126
Db 715 CCCGGAGCAGCAGGCGCGCGCGCCGCCGCCGCCCGCTGGAGCGAGCCGAGGC 774
QY 127 GGCCTGTGCGCCGAGTGGTGTGCTCTTGACCATGCGCGAGCGCTTCATTGCAAGCCG 186
Db 775 GGCATTGGAGCAGGAAACCGGCGCGCGCGCGCGAGCGAGCTTGCTTCAGCAAAAGGG 834
QY 187 CGAGAGCTCAGCTCCGACCCCGCGCGCGCGCTTCCAGAGAGGTTTCATCTGCTCT 246
Db 835 TCGCGAGGCGCGGAGCGCGCGAGCTCGAGTCCCGCAGAGGAAAGCAAGTGCACATA 894
QY 247 CCCGTGCTGCGAGCGCCCAATTACAGAAAGCCTGGAAGCTTGACCGGACGCTGTCAGCA 306
Db 895 GCGGCGCTGCGAGAAAGTTTACGGAATCTTCCACCTCAAGGCGACCTGGAAGACTCA 954
QY 307 CACGGGGAGAGACCATTTGTTGTGACTATGAAGGCTGTGCAAGGCTTCATCAGGGA 366
Db 955 CACAGGTGAGAGGCTTTGCGCTGCAGCTGGCAGAGTGCAGCAAGATTCGACGCTC 1014
QY 367 CTACCATCTGAGCCGCACTTGTGACTACACAGAGGAAAGCCGTT 414
Db 1015 GGACGAGCTGGCAGGCACTATGCGACGACACGCGCGAGAAAGATT 1062

RESULT 4
US-09-492-985-1
Sequence 1, Application US/09492985
Patent No. 6376240
GENERAL INFORMATION:
APPLICANT: Song, An M.
APPLICANT: Chen, Ya-Fen
TITLE OF INVENTION: REFAT-1: A Transcription Factor That
FILE REFERENCE: SUN-113P
CURRENT APPLICATION NUMBER: US/09/492,985
EARLIER FILING DATE: 2000-01-27
EARLIER APPLICATION NUMBER: 60/117,576
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1430
TYPE: DNA
ORGANISM: homo sapien
FEATURE:
OTHER INFORMATION: synthetic oligonucleotide probe

OTHER INFORMATION: synthetic oligonucleotide probe
US-09-492-985-1

Query Match 5.1%; Score 61.4; DB 4; Length 1430;
Best Local Similarity 48.9%; Pred. No. 6e-07;

Matches 193; Conservative 0; Mismatches 201; Indels 1; Gaps 1;

```
OY 21 TTCACGAGGAGCCGTGGCCGCGCGCGGTCCTCCCGGACGTCGTCCGACGTGCA 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 664 TCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
OY 81 GCGGCGCTGGCCCTGGCTTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 724 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
OY 140 GAGTCGTCGTCCTGTCGACATCGCGGCGCGTCCTTCATTCGACGCGCGGAGCTCAGCT 199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 784 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 843
OY 200 CCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 844 AGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
OY 260 GCCATTACAGCAAGCCTGGAGCTTGGAGCGCGCGCGCGCGCGCGCGCGCGCG 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 904 AAGCTTACGCGCAATCTTGGCAGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCG 963
OY 320 CCATTTGTTGTCATGTAAGAGGTGGCAAGCGCTTCATCAGGACTACCATCTGAGC 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 964 CCCTTCGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1023
OY 380 CGCCACATTCGTGACTCAGACAGAGAAAGCCGTT 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1024 CGGCACTACGCGCACACAGCGCGCGAGAAAGATT 1058
```

RESULT 5

US-09-362-123A-3

```
; Sequence 3, Application US/09362123A
; Patent No. 6451558
; GENERAL INFORMATION:
; APPLICANT: Cooke, Michael Paul
; APPLICANT: Holness, Claire
; APPLICANT: Sirenko, Oksana
; TITLE OF INVENTION: No. 6451558e1 Genes in the Control of Hematopoiesis
; FILE REFERENCE: 4-30629A/SYS
; CURRENT APPLICATION NUMBER: US/09/362.123A
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: 09/128.310
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2992
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82)..(1710)
US-09-362-123A-3
```

Query Match 4.8%; Score 58.8; DB 4; Length 2992;
Best Local Similarity 54.6%; Pred. No. 4.1e-06;

Matches 142; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

```
OY 178 TGCAGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCTTCC 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1263 TGCCTTACGCGCGCGCTCTCTTTAAGCAACAGCGCGGTGCACACAGCGGAGAGCC 1322
OY 238 CTGCTGCTCCCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1323 CTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1382
```

```
OY 298 GTGCAGACACAGCGGAGAGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGGCTT 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1383 GCTTCCTGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1436
OY 358 CATCAGGAGCTACCATTTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1437 CGCCAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1496
OY 418 TTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1497 GTGTACGCGAGTGTGCGCGCG 1516
```

RESULT 6

US-09-234-332-5

```
; Sequence 5, Application US/09234332A
; Patent No. 6087168
; GENERAL INFORMATION:
; APPLICANT: Cedars-Sinai Medical Center
; APPLICANT: Michael F. Levesque, M.D.
; APPLICANT: Thomas Neuman, Ph.D.
; TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
; FILE REFERENCE: P07 41494
; CURRENT APPLICATION NUMBER: US/09/234.332A
; CURRENT FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3138
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: gene
; LOCATION: (0)..(0)
OTHER INFORMATION: Zlc 1 Protein gene; Genbank Accession D76435
US-09-234-332-5
```

Query Match

Best Local Similarity 55.2%; Score 54.8; DB 3; Length 3138;
Matches 107; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
OY 281 AAGCTGACGCGCACCTGTGCAAGCACAGCGGAGAGACCATTTGTTGTGACTATGAA 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1642 AAATCGTTAAACACATCCGCGTGCACAGCGGAGAAAGCCCTTCCCTCCCTCCCT 1701
OY 341 GGGTGTGCAAGCGCTTATCAGGAGCTACCATCTGAGCGCGCGCGCGCGCGCGCG 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1702 GCGTGTGCAAGCGCTTCCGCGCGCTCCGAGAAATTAAAGATCCACAAAGAGCGCAC 1761
OY 401 GGAAGAAAGCGTTGTTGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1762 GGGGAGAAAGCGCTTCAAGTGGAGTTGAGGGCTGTGACCGCGCGCTTCCGTAACAGCAGC 1821
OY 461 AACTGAAGAAACA 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1822 GACCGCAAGAAACA 1835
```

RESULT 7

US-08-570-227A-1

```
; Sequence 1, Application US/08570227A
; Patent No. 5981217
; GENERAL INFORMATION:
; APPLICANT: Subramaniam, M.
; APPLICANT: Spelsberg, T. C.
; TITLE OF INVENTION: DNA ENCODING TGF-BETA INDUCIBLE
; TITLE OF INVENTION: EARLY FACTOR-1 (TIEF-1), A GENE EXPRESSED
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Schweigman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
```

```

: CITY: Minneapolis
: STATE: MN
: COUNTRY: USA
: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/570,227A
: FILING DATE: 11-DEC-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Woessner, Warren D
: REGISTRATION NUMBER: 30,440
: REFERENCE/DOCKET NUMBER: 150,157US1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-359-3260
: TELEFAX: 612-359-3263
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2881 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-570-227A-1

```

```

Query Match          4.4%: Score 53.2; DB 2; Length 2881;
Best Local Similarity 54.6%: Pred. No. 0.00013;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

QY 227 AGGAGTTTCATCTCTCTCTCTCTGACGACGCGCAATTACAGCAAAAGCCTGGAAGCTT 286
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1185 AGGAGTTCATCTCTGACGACGCGCAAGATGTGGCAAGACATCTTAAAGTTCCCATCTG 1244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 GACGCGCACCTGTGCAAGCAGCGGGGAGAGACCATTTGTTGTGACTATGAAGGGTGT 346
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1245 AAGGCCACAGAGGAGAGCAGACAGAGAAAAGCCTTTCAGCTGAGCTGGAAGGGTTGT 1304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 GCGAAGCCTTCATCAGGAGATACCATCTGAGCCGCCACATCTGACTACACAGAGAGAA 406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1305 GAAAGGAGCTTGGCCGCTTGTGATGAAGCTGTCCAGACAGCGAAGCCACAGCGGTGAG 1364
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 AAGCCGTTTGTGTTG 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1365 AAGAAATTTGGCGTG 1378
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 8
US-09-077-991-1
: Sequence 1, Application US/09077991
: Patent No. 6207375
: GENERAL INFORMATION:
: APPLICANT: Subramaniam, M.
: APPLICANT: Speisberg, T. C.
: APPLICANT: Roche, P. C.
: TITLE OF INVENTION: TGF-beta inducible early factor-1
: TITLE OF INVENTION: (TIEF-1) and a method to detect breast cancer
: FILE REFERENCE: 150,157052
: CURRENT APPLICATION NUMBER: US/09/077,991
: EARLIER FILING DATE: 1998-07-07
: EARLIER APPLICATION NUMBER: PCT/US96/19555
: EARLIER FILING DATE: 1996-12-11
: EARLIER APPLICATION NUMBER: US 08/570,227
: EARLIER FILING DATE: 1995-12-11
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: FastSeq for Windows Version 3.0

```

```

: SEQ ID NO 1
: LENGTH: 2881
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-077-991-1

```

```

Query Match          4.4%: Score 53.2; DB 4; Length 2881;
Best Local Similarity 54.6%: Pred. No. 0.00013;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

QY 227 AGGAGTTTCATCTCTCTCTCTGACTGACGCGCAATTACAGCAAAAGCCTGGAAGCTT 286
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1185 AGGAGTTCATCTGACGACGCGCAAGATGTGGCAAGACATCTTAAAGTTCCCATCTG 1244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 GACGCGCACCTGTGCAAGCAGCGGGGAGAGACCATTTGTTGTGACTATGAAGGGTGT 346
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1245 AAGGCCACAGAGGAGAGCAGACAGAGAAAAGCCTTTCAGCTGAGCTGGAAGGGTTGT 1304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 GCGAAGCCTTCATCAGGAGATACCATCTGAGCCGCCACATCTGACTACACAGAGAGAA 406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1305 GAAAGGAGCTTGGCCGCTTGTGATGAAGCTGTCCAGACAGCGAAGCCACAGCGGTGAG 1364
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 AAGCCGTTTGTGTTG 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1365 AAGAAATTTGGCGTG 1378
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 9
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300,6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:

```



```
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/946,241B
? FILING DATE: 07-OCT-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/030,035
? FILING DATE: 05-NOV-1996
? APPLICATION NUMBER: 60/027,521
? FILING DATE: 07-OCT-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Creason, Gary L.
? REGISTRATION NUMBER: 34,310
? REFERENCE/DOCKET NUMBER: 05433/027001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617-542-5070
? TELEFAX: 617-542-8906
? TELEX: 200154
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1889 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 434...1843
? US-08-946-241B-1
```

```
Query Match          4.2%; Score 50.4; DB 2; Length 1889;
Best Local Similarity 55.0%; Pred. No. 0.00059;
Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
OY 235 CATTCGCTCCTCCCTGACAGCGCCCAATTACAGCAAAAGCTTGAGCGGCA 294
      || || || || || || || || || || || || || || || || || ||
DB 1594 CACTTGTGATTACCGGGGCGGCGCAAAACCTACAAAGATTCCCATCTCAAGGCACA 1653

OY 295 CCTGTGACAGACACAGGGGGAGAGACCATTTGTTGTGACTATGAAGGTGTGGCAAGGC 354
      || || || || || || || || || || || || || || || || || ||
DB 1654 CCTGCGAACCACACAGAGTGAGAAACCTTACCACTGTGACTGGGACGGCTGTGATGAA 1713

OY 355 CTTTCATGAGGACTACATCTGAGCGCCACATTTCTGACTCACAGAGAGAAAAGCCGTT 414
      || || || || || || || || || || || || || || || || || ||
DB 1714 ATTGCGCCGCTCAGATGACTGACGACGAGCACTACCGTAACACACAGGGGACACGCCCGTT 1773
```

```
RESULT 13
US-08-946-241B-8
; Sequence 8, Application US/08946241B
; Patent No. 5928941
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: MCA Nulty, Megan M.
; TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,241B
; FILING DATE: 07-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 60/030,035
```

```
? FILING DATE: 05-NOV-1996
? APPLICATION NUMBER: 60/027,521
? FILING DATE: 07-OCT-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Creason, Gary L.
? REGISTRATION NUMBER: 34,310
? REFERENCE/DOCKET NUMBER: 05433/027001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617-542-5070
? TELEFAX: 617-542-8906
? TELEX: 200154
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1889 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 407...1843
? US-08-946-241B-8
```

```
Query Match          4.2%; Score 50.4; DB 2; Length 1889;
Best Local Similarity 55.0%; Pred. No. 0.00059;
Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
OY 235 CATTCGCTCCTCCCTGACAGCGCCCAATTACAGCAAAAGCTTGAGCGGCA 294
      || || || || || || || || || || || || || || || || || ||
DB 1594 CACTTGTGATTACCGGGGCGGCGCAAAACCTACAAAGATTCCCATCTCAAGGCACA 1653

OY 295 CCTGTGACAGACACAGGGGGAGAGACCATTTGTTGTGACTATGAAGGTGTGGCAAGGC 354
      || || || || || || || || || || || || || || || || || ||
DB 1654 CCTGCGAACCACACAGAGTGAGAAACCTTACCACTGTGACTGGGACGGCTGTGATGAA 1713

OY 355 CTTTCATGAGGACTACATCTGAGCGCCACATTTCTGACTCACAGAGAGAAAAGCCGTT 414
      || || || || || || || || || || || || || || || || || ||
DB 1714 ATTGCGCCGCTCAGATGACTGACGACGAGCACTACCGTAACACACAGGGGACACGCCCGTT 1773
```

```
RESULT 14
US-09-309-053-1
; Sequence 1, Application US/09309053
; Patent No. 6077933
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: MCA Nulty, Megan M.
; TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,053
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/946,241
; FILING DATE: 07-OCT-1997
; APPLICATION NUMBER: 60/030,035
; FILING DATE: 05-NOV-1996
; APPLICATION NUMBER: 60/027,521
; FILING DATE: 07-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
```

```

REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 434...1843
US-09-309-053-1

Query Match          4.2%   Score 50.4; DB 3; Length 1889;
Best Local Similarity 55.0%; P: 0.00059;
Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY    235 CATCTGCTCTCCCTCGATCGGCACCAATTATCAGCAAGCGCTGGAGGCTTGACGCCGA 294
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1554 CACTTGTATTAACGGGGCGCTGGGGAAAAAACCCTACCAAAGAATTCCCATTCGAAGCACA 1653

OY    295 CCTGTGCAAGCACACAGGGGGAGAGACCATTTGTTTTGTGACTATGAAGGTGTGGCAAGCC 354
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1654 CCTCGAACCCACACAGGTGAGAAACCTTACCACTGTGACTCGGACCGCTGTGATGGAAA 1713

OY    355 CTTCATCAGGAGCTACCATCTGCAGCCGCCACACTTTCTGACTCACAACAGAGAAGAACCGCTT 414
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1714 ATTCCCGCGCTTCAGTAGTAACCTGACAGGCACCTACCGTAAACACAGGGGACCGCGCGCTT 1773
```

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
US-09-309-053-8

Query Match          4.2%: Score 50.4; DB 3; Length 1889;
Best Local Similarity 55.0%: Pred. No. 0.00059;
Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 235 CATCTGCTCTTCCTCCCTGACTGCAGCGGCCCATTTACAGCAAGAGCCCTGGAAGCTTGACCGCGCA 294
Db 1594 CACCTTGATATTATGAGGGGGCTGCGGCAAAACCTACACAAAGAGTCTCCCATCTAAAGGCACA 1653

OY 295 CCTGTCAAGCAGACAGGGGGAGAGACCAATTTGTGTGACTATGAAGGTTGTGGCAAGGC 354
Db 1654 CCTGGCAACCCACAGAGGTGAGAAACCTTTAGCACGTGTGACTGGAGACGGCTGTGCATGGAA 1713

OY 355 CTTTCATCAGAGGACTACCACTGTGAGCGGCCACACTTCTGACTGCACAGCAGAAAGCCGTTT 414
Db 1714 ATTCCCGCGCTCAGATGAAGACTGACCAGGCACACTAACCGTATTAACACAGGGGCACCGCCGTT 1773

Search completed: February 10, 2003, 12:48:08
Job time : 109.257 secs

```

Search completed: February 10, 2003, 12:48:08
Job time : 109.257 secs

RESULT 15
 US-09-309-053-8
 Sequence 8, Application US/09309053
 Patent No. 6077933
 GENERAL INFORMATION:
 APPLICANT: Lee, Mu-En
 APPLICANT: MCA'Nulty, Megan M.
 TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/309,053
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/946,241
 FILING DATE: 07-OCT-1997
 APPLICATION NUMBER: 60/030,035
 FILING DATE: 05-NOV-1996
 APPLICATION NUMBER: 60/027,521
 FILING DATE: 07-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Creason, Gary L.
 REGISTRATION NUMBER: 34,310
 REFERENCE/DOCKET NUMBER: 05433/027001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 TELETYPE: 200154
 INFORMATION FOR SEQ ID NO: 8:

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 1889 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 407...1843
: US-09-309-053-8

Query Match      4.2%; Score 50.4; DB 3; Length 1889;
Best Local Similarity 55.0%; Pred. No. 0.00059;
Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 235 CATCTGCTCTCTCCCTGACTGCGAGGCCAATTACAGCAAAAGCCTTGAAAGCTTGACGCGCA 294
    || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1594 CACTTGTGATTACGGGGGCTGCGGCAAAACCTACACAAGAGTTCCCATCTCAAGGCACA 1653

OY 295 CCTGCAAGCACACAGGGGGAGAGACCAATTGTTGTGACTATGAAAGGCTGCGCAAGGC 354
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1654 CCTGCAAGCCACACAGGAGAGAAACCTTACCACCTGTGACTGGAACGGCTGTGATGAA 1713

OY 355 CTTTCATCAGAGGACTACCACTGTGAGCCGACACATTCGACTGCACAGCAGAAAGCCGTT 414
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1714 ATTGCGCGGCTCAGATGAACTGACGACGACACTACCGTAAACACAGGGGCACCGCCGTT 1773

Search completed: February 10, 2003, 12:48:08
Job time : 109.257 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:38:15 : Search time 64.895 Seconds
(without alignments)
8859.935 Million cell updates/sec

Title: US-09-831-426c-4
Perfect score: 1213
Sequence: 1 gtgcgcgcgcgcgcgaag.....cagtaactacccttgctaa 1213

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 408267 seqs, 237001491 residues

Total number of hits satisfying chosen parameters: 816534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	418.4	34.5	439	10 US-09-920-300A-326	Sequence 326, App
C 2	418.4	34.5	439	12 US-10-033-528-326	Sequence 326, App
C 3	416.8	34.4	439	10 US-09-920-300A-689	Sequence 689, App
C 4	416.8	34.4	439	12 US-10-033-528-689	Sequence 689, App
C 5	234	19.3	234	9 US-09-796-692-406A	Sequence 406A, Ap
C 6	234	19.3	234	10 US-09-920-300A-610	Sequence 610, App
C 7	234	19.3	234	12 US-10-033-528-610	Sequence 610, App
C 8	226	18.6	947	9 US-09-920-758-68	Sequence 68, Appl
C 9	152.4	12.6	947	9 US-09-920-758-68	Sequence 68, Appl
C 10	152.4	12.6	947	10 US-09-920-758-68	Sequence 68, Appl
C 11	152.4	12.6	947	10 US-09-920-758-68	Sequence 68, Appl
C 12	152.4	12.6	947	10 US-09-920-758-68	Sequence 68, Appl
C 13	152.4	12.6	947	10 US-09-920-758-68	Sequence 68, Appl
C 14	152.4	12.6	947	10 US-09-920-758-68	Sequence 68, Appl
C 15	81	6.7	368	10 US-09-920-300A-1747	Sequence 1747, Ap
C 16	81	6.7	368	12 US-10-033-528-1747	Sequence 1747, Ap
C 17	80.4	6.6	2400	10 US-09-864-761-18874	Sequence 18874, A
C 18	69.2	5.7	2960	9 US-09-734-329-1	Sequence 1, Appl1
C 19	64.6	5.3	292	9 US-09-942-090-14	Sequence 14, Appl

20	64.6	5.3	292	10 US-09-925-796-14	Sequence 14, Appl
21	64.6	5.3	292	10 US-09-941-450-14	Sequence 14, Appl
22	63	5.2	298	9 US-09-942-087A-14	Sequence 14, Appl
23	63	5.2	298	10 US-09-779-233-2	Sequence 2, Appl1
24	63	5.2	298	10 US-09-844-508-10	Sequence 10, Appl
25	62	5.1	3213	12 US-10-002-600-76	Sequence 4943, Ap
26	61.8	5.1	547	10 US-09-867-701-4943	Sequence 1647, Ap
27	61.8	5.1	3197	10 US-09-764-864-1647	Sequence 29, Appl
28	60.6	5.0	589	9 US-09-942-087A-29	Sequence 17, Appl
29	60.6	5.0	589	10 US-09-779-233-17	Sequence 28, Appl
30	60.6	5.0	589	10 US-09-844-508-28	Sequence 211, App
31	60.2	5.0	900	10 US-09-764-864-211	Sequence 417, App
32	60.2	5.0	1602	10 US-09-833-790-417	Sequence 4925, Ap
33	60	4.9	259	10 US-09-960-352-4925	Sequence 642, App
34	60	4.9	882	10 US-09-764-864-642	Sequence 16, Appl
35	58.2	4.8	298	9 US-09-942-087A-16	Sequence 4, Appl1
36	58.2	4.8	298	10 US-09-779-233-4	Sequence 17, Appl
37	58.2	4.8	298	10 US-09-844-508-17	Sequence 950, App
38	55.4	4.6	1301	9 US-09-954-531-950	Sequence 1613, Ap
39	55.4	4.6	1301	10 US-09-880-107-1613	Sequence 120, App
40	55.4	4.6	1301	10 US-09-967-768A-120	Sequence 73, Appl
41	55.4	4.6	2572	10 US-09-919-172-73	Sequence 137, App
42	55.4	4.6	3332	9 US-10-071-766-137	Sequence 1716, Ap
43	54.8	4.5	3138	10 US-09-880-107-1716	Sequence 119, App
44	54.6	4.5	396	9 US-09-970-966-119	Sequence 119, App
45	54.6	4.5	396	10 US-09-825-294-119	

ALIGNMENTS

RESULT 1
US-09-920-300A-326/c
; Sequence 326, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 326
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-326

Query Match 34.5%: Score 418.4: DB 10: Length 439:

Best Local Similarity 99.8%: Pred. No. 1.4e-103: Matches 419: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY	794	AAACGCAAGATTACCTTAAGCAACACATGAAATCTATGCCGAGAAAGGATGTGT	853
DB	439	AAACGCAAGATTACCTTAAGCAACACATGAAATCTATGCCGAGAAAGGATGTGT	380
QY	854	CGCGTCCAGAGAGAGCGCTGGAGAAACATCTCTGTTTATTCGCAAGGCAT	913
DB	379	CGCGTCCAGAGAGAGCGCTGGAGAAACATCTCTGTTTATTCGCAAGGCAT	320
QY	914	ATCCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGCTGCGTGGCAAA	973
DB	319	ATCCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGCTGCGTGGCAAA	260
QY	974	ACATTTCGATGAAAGGATCTACTAGGATCTGTTGTACTGATCTGACAAAG	1033
DB	259	ACATTTCGATGAAAGGATCTACTAGGATCTGTTGTACTGATCTGACAAAG	200

QY	1034	AAATGGAAGTC	CAAAAGTCA	AAATATCGT	GGAAAAAGG	AGTTTGCC	CTCATTCT	AGT	1093
Db	199	AAAAATGAAGT	CAAAAGTCA	AAAAATATCGT	GGAAAAAGG	AGTTTGCC	CTCATTCT	AGT	140
QY	1094	GGATATATTC	CCCTCCCA	AAAGCAAGG	CAAGCTTAT	CTTGTGTG	CAAAACG	AGAG	1153
Db	139	GGATATATTC	CCCTCCCA	AAAGCAAGG	CAAGCTTAT	CTTGTGTG	CAAAACG	AGAG	80
QY	1154	TCACCCAACT	GTGTGGA	AGACAGAT	CTCTGCAC	AGTTGCA	GTACTTAC	CCCTTG	GGCTAA
Db	79	TCACCCAACT	GTGTGGA	AGACAGAT	CTCTGCAC	AGTTGCA	GTACTTAC	CCCTTG	GGCTAA

```

RESULT 2
US-10-033-528-326/c
: Sequence 326, Application US/10033528
: Patent No. US20020131971A1
: GENERAL INFORMATION:
: APPLICANT: Kling, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547c1
: CURRENT APPLICATION NUMBER: US/10/033,528
: CURRENT FILING DATE: 2001-12-26
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 326
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-033-528-326

```

Query Match	34.5%	Score 418.4	DB 12	Length 439
Best Local Similarity	99.8%	Pred No. 1.4e-103		
Matches 419; Conservative	0	Mismatches 1	Indels 0	Gaps 0

OY	794	AAAGCAAAAGTTATTCCTTTAAGCAACACATGAAAACCTCATGCCCGCAAAAAGGATATGAT	853
Db	439	AAAGCAAAAGTTATTCCTTTAAGCAACACATGAAAACCTCATGCCCGCAAAAAGGATATGAT	380
OY	854	CGCTGTCCAAAGAGAAGGCTGTGGAAAGAACCTAATCTACTGTGTATATCTCCAAAGCCAT	913
Db	379	CGCTGTCCAAAGAGAAGGCTGTGGAAAGAACCTAATCTACTGTGTATATCTCCAAAGCCAT	320
OY	914	ATCTCTTCCTTTCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGTGCTGTGGCAAA	973
Db	319	ATCTCTTCCTTTCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGTGCTGTGGCAAA	260
OY	974	ACATTTGCAATGAACAAACATCTCAGTACGATCGTGTGACATGATCTGTACAAAG	1033
Db	259	ACATTTGCAATGAACAAACATCTCAGTACGATCGTGTGACATGATCTGTACAAAG	200
OY	1034	AAATGGAAGCTCAAAAGTCAAAAATATCTCGGAAAGAAAGGAGTTTGCCCTCATCTCACT	1093
Db	199	AAATGGAAGCTCAAAAGTCAAAAATATCTCGGAAAGAAAGGAGTTTGCCCTCATCTCACT	140
OY	1094	CGATATATTCCTCCCAAAAGGAAACAAGGCAAGGCTTATCTTTGTGTCAAAAGCGAGAG	1153
Db	139	CGATATATTCCTCCCAAAAGGAAACAAGGCAAGGCTTATCTTTGTGTCAAAAGCGAGAG	80
OY	1154	TCACCCAAAGCTGTGGGAAGACAAGATGTCTGTGACAGATTGCAAGTACTTACCTTGGCTAA	1213
Db	79	TCACCCAAAGCTGTGGGAAGACAAGATGTCTGTGACAGATTGCAAGTACTTACCTTGGCTAA	20

RESULT 3
US-09-920-300A-689
; Sequence 689, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:

```

APPLICANT: King, Gordon E.
APPLICANT: Meagher, Madeline Joy
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.547
CURRENT APPLICATION NUMBER: US/09/920,300A
CURRENT FILING DATE: 2001-07-31
NUMBER OF SEQ ID NOS: 1789
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 689
LENGTH: 439
TYPE: DNA
ORGANISM: Homo sapiens
US-09-920-300A-689

Query Match          34.4%   Score 416.8; DB 10; Length 439;
Best Local Similarity 99.5%;   Pred. No. 3.7e-103;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0

```

OY	794	AAAGCGAAAGTTTACCTTTAAGCAAAACATGAAAAATCATGCCCGCAAGAAAGGATATATG	853
Db	1	AAAGCGAAAGTTTACCTTTAAGCAAAACATGAAAAATCATGCCCGCAAGAAAGGATATATG	60
OY	854	CGCTGTCCAGAGAAAGGCTGTGGAAAGACCTATACTCTGTGTTTAATCTCCAAAGCCAT	913
Db	61	CGCTGTCCAGAGAAAGGCTGTGGAAAGACCTATACAACTGTGTTTAATCTCCAAAGCCAT	120
OY	914	ATCTCTCTCTTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAAA	973
Db	121	ATCTCTCTCTTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAAA	180
OY	974	ACATTTGCATGAAACAACACTCTCATTAGGCAATGCTGTGATCATGATCTGTACAAGAAAG	1033
Db	181	ACATTTGCATGAAACAACACTCTCATTAGGCAATGCTGTGTAACATGATCTGTACAAGAAAG	240
OY	1034	AAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAAAAAAGGAGTTTGGCTCTCATCTCAGT	1093
Db	241	AAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAAAAAAGGAGTTTGGCTCTCATCTCAGT	300
OY	1094	GGATATATTCCTCCCAAAAGGAAACAAGGCAAGGCTTATCTTTGTGTCAAAACGGAGAG	1155
Db	301	GGATATATTCCTCCCAAAAGGAAACAAGGCAAGGCTTATCTTTGTGTCAAAACGGAGAG	360
OY	1154	TCACCCCAAGCTGTGGAAAGCAAGATGCTGTGACAGTTGGAGTACTTACCTTGGGCTAA	1213
Db	361	TCACCCCAAGCTGTGGAAAGCAAGATGCTGTGACAGTTGGAGTACTTACCTTGGGCTAA	420

```

Query Match      34.4%; Score 416.8; DB 13; Length 439
US-10-033-528-689
RESULT 4
: Sequence 689, Application US/10033528
: Patent No. US20020131971A1
: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jianshun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547C1
: CURRENT APPLICATION NUMBER: US/10/033,528
: CURRENT FILING DATE: 2001-12-26
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 689
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-033-528-689

```



```
Best Local Similarity 99.5%; Pred. No. 3,7e-103;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 794 AAAGCAAGATTCCTTAAGCAACATGAAACATGCCCCAGAAAGGATGTATGT 853
    |||||||
Db 1 AAAGCAAGATTCCTTAAGCAACATGAAACATGCCCCAGAAAGGATGTATGT 60
    |||||||
QY 854 CGCTGCCAAGAGAGCGCTGTGAGAACCTATCTAGTGTGTTAATCTCAAGCAT 913
    |||||||
Db 61 CGCTGCCAAGAGAGCGCTGTGAGAACCTATCTAGTGTGTTAATCTCAAGCAT 120
    |||||||
QY 914 ATCTCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTAACATGCTGGCTGTGGCAA 973
    |||||||
Db 121 ATCTCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTAACATGCTGGCTGTGGCAA 180
    |||||||
QY 974 ACATTGCAATGAACAAGCTCAGTACGATGCTGTGACATGATCTCGAACAAG 1033
    |||||||
Db 181 ACATTGCAATGAACAAGCTCAGTACGATGCTGTGACATGATCTCGAACAAG 240
    |||||||
QY 1034 AAATGAAGCTCAAAAGTCAAAATCTGTCGAAAAAGGAGATTGGCTCTCATCTCAGT 1093
    |||||||
Db 241 AAATGAAGCTCAAAAGTCAAAATCTGTCGAAAAAGGAGATTGGCTCTCATCTCAGT 300
    |||||||
QY 1094 GGATATATCCCTCCCAAAAGCAAGGCGAAGGCTTATCTTTGTGTCAAAACGAGAG 1153
    |||||||
Db 301 GGATATATCCCTCCCAAAAGCAAGGCGAAGGCTTATCTTTGTGTCAAAACGAGAG 360
    |||||||
QY 1154 TCACCCCAACTGTGTGGAAGACAAGATGCTGCGACAGTGCAGTACCTGGGCTAA 1213
    |||||||
Db 361 TCACCCCAACTGTGTGGAAGACAAGATGCTGCGACAGTGCAGTACCTGGGCTAA 420
    |||||||
```

```
RESULT 5
US-09-796-692-4064
; Sequence 4064, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
```

```
; APPLICANT: Mannion, Jane
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
```

```
; FILE REFERENCE: 2077, 001200
```

```
; CURRENT APPLICATION NUMBER: US/09/796,692
```

```
; PRIOR FILING DATE: 2001-03-01
```

```
; PRIOR APPLICATION NUMBER: 60/186,126
```

```
; PRIOR FILING DATE: 2000-03-01
```

```
; PRIOR APPLICATION NUMBER: 60/190,479
```

```
; PRIOR FILING DATE: 2000-03-17
```

```
; PRIOR APPLICATION NUMBER: 60/200,545
```

```
; PRIOR FILING DATE: 2000-04-27
```

```
; PRIOR APPLICATION NUMBER: 60/200,303
```

```
; PRIOR FILING DATE: 2000-04-28
```

```
; PRIOR APPLICATION NUMBER: 60/200,779
```

```
; PRIOR FILING DATE: 2000-04-28
```

```
; PRIOR APPLICATION NUMBER: 60/200,999
```

```
; PRIOR FILING DATE: 2000-05-01
```

```
; PRIOR APPLICATION NUMBER: 60/202,084
```

```
; PRIOR FILING DATE: 2000-05-04
```

```
; PRIOR APPLICATION NUMBER: 60/206,201
```

```
; PRIOR FILING DATE: 2000-05-22
```

```
; PRIOR APPLICATION NUMBER: 60/218,950
```

```
; PRIOR FILING DATE: 2000-07-14
```

```
; PRIOR APPLICATION NUMBER: 60/222,903
```

```
; PRIOR FILING DATE: 2000-08-03
```

```
; PRIOR APPLICATION NUMBER: 60/223,416
```

```
; PRIOR FILING DATE: 2000-08-04
```

```
; PRIOR APPLICATION NUMBER: 60/223,378
```

```
; PRIOR FILING DATE: 2000-08-07
```

```
; NUMBER OF SEQ ID NOS: 9597
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 4064
```

```
; LENGTH: 234
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4064
```

```
Query Match 19.3%; Score 234; DB 9; Length 234;
Best Local Similarity 100.0%; Pred. No. 7.1e-54;
```

```
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 560 CTGAAATCCATGCTGCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGA 619
    |||||||
Db 1 CTGAAATCCATGCTGCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGA 60
    |||||||
QY 620 TGTGGAAACACTTGTGCATCACCAGCAAGCTGAAAGACATGCAAGGCCAGAGGGC 679
    |||||||
Db 61 TGTGGAAACACTTGTGCATCACCAGCAAGCTGAAAGACATGCAAGGCCAGAGGGC 120
    |||||||
QY 680 TATGTATGTCAAAAAGAGATGCTCTTTGTGGCAAAAACATGAGAGCAACTTGTGAACAT 739
    |||||||
Db 121 TATGTATGTCAAAAAGAGATGCTCTTTGTGGCAAAAACATGAGAGCAACTTGTGAACAT 180
    |||||||
QY 740 GTGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGCGGGAACATTT 793
    |||||||
Db 181 GTGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGCGGGAACATTT 234
    |||||||
```

```
RESULT 6
US-09-920-300A-610
```

```
; Sequence 610, Application US/09920300A
```

```
; Patent No. US20020136728A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: King, Gordon E.
```

```
; APPLICANT: Meagher, Madeleine Joy
```

```
; APPLICANT: Xu, Jiangchun
```

```
; APPLICANT: Secrist, Heather
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
```

```
; FILE REFERENCE: 210121,547
```

```
; CURRENT APPLICATION NUMBER: US/09/920,300A
```

```
; CURRENT FILING DATE: 2001-07-31
```

```
; NUMBER OF SEQ ID NOS: 1789
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 610
```

```
; LENGTH: 234
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
US-09-920-300A-610
```

```
Query Match 19.3%; Score 234; DB 10; Length 234;
```

```
Best Local Similarity 100.0%; Pred. No. 7.1e-54;
```

```
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 560 CTGAAATCCATGCTGCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGA 619
    |||||||
Db 1 CTGAAATCCATGCTGCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGA 60
    |||||||
QY 620 TGTGGAAACACTTGTGCATCACCAGCAAGCTGAAAGACATGCAAGGCCAGAGGGC 679
    |||||||
Db 61 TGTGGAAACACTTGTGCATCACCAGCAAGCTGAAAGACATGCAAGGCCAGAGGGC 120
    |||||||
QY 680 TATGTATGTCAAAAAGAGATGCTCTTTGTGGCAAAAACATGAGAGCAACTTGTGAACAT 739
    |||||||
Db 121 TATGTATGTCAAAAAGAGATGCTCTTTGTGGCAAAAACATGAGAGCAACTTGTGAACAT 180
    |||||||
QY 740 GTGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGCGGGAACATTT 793
    |||||||
Db 181 GTGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGCGGGAACATTT 234
    |||||||
```

```
RESULT 7
```

```
US-10-033-528-610
```

```
; Sequence 610, Application US/10033528
```

```
; Patent No. US20020131971A1
```

```
; GENERAL INFORMATION:
```



```
; Sequence 5, Application US/09732348
; Patent No. US20020046419A1
; GENERAL INFORMATION:
; APPLICANT: Yen Choo, et al.
; TITLE OF INVENTION: Regulated Gene Expression in Plants
; FILE REFERENCE: 674538-2001
; CURRENT APPLICATION NUMBER: US/09/732.348
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 947
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (773)..(908)
; OTHER INFORMATION: transactivation domain of VP64, other features except c-myc tag
; OTHER INFORMATION: listed below) same as SEQ ID NO:
; NAME/KEY: misc_feature
; LOCATION: (909)..(938)
; OTHER INFORMATION: c-myc tag, other features except transactivation domain VP64 (115
; OTHER INFORMATION: ted above) same as SEQ ID NO:
US-09-732-348-5

Query Match          12.6%; Score 152.4; DB 10; Length 947;
Best Local Similarity 57.6%; Pred. No. 1,9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 208 GCCGCGCCCGCGCTCCCGAGAGGTCATCTGCTCTCCCTGACGCGCCCAATTA 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26  GCGCGTGGCGGTGGTATAGCGGTACATCTGCTCTTCCCGACGCGCGCTTA 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 268 CAGCAAGCCCTGGAACTGTGACGGGACCTGTGCAAGCAGCGGGGAGAACCATTTGT 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86  TAAACAAGAACTGAAACTGCAAGCGCATCTGTGCAAAACACAGAGAGAAACCATTTCC 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 328 TTGTGACTATGAAGGTTGTGCGAGCGCTTCATCAGGAGTACCATGAGCGGCACAT 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146  ATGTAGAGAGAGAGAGATGTGAGAAAGCTTTACCTCGCTTATACCTTAACCGGCACATC 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 388 TCTGACTCACACAGAGAGAAAGCGTTTGTGTGACAGCCACTGGCTGTGATCAAAAT 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206  ACTGACTCATCTGCGCGAGAAACCTTCACATGTGACGTGGATGATGTGATGAGATT 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 448 CAACACAATCAAACTTGAAGAAACATTTTGAACGCAACATGAAATCAACAAAACA 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266  TACTACAAAGGCAACATGAAGACACTTACAGATTTCATTAACATCAAGATCTGCGT 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 508 ATATATATGAGTTTGAAGACTGTAAAGACCTTTAAGAAACATGACAGCTGAAAT 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326  CTATGTGTGCCATTTTGAAGACTGTGCAAGCATTCAGAGAACATCAATTAAGGT 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 568 CCATCAGTCCAGCATACCAATGAACCTTATTCAGTGTACCCAGAGAGATGTGGAA 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386  TCATCAGTTCAGTCAACACAGACAGCTGCCGTATGCTTGCCTGTGCGATCG 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 628 ACACATTGATCATCCAGCAAGCTGAAAGCATGCCAAGGCCCAACAGGGCTA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446  CCGGTTTTCGTGCTGGATGAGCTTACCCGCCATATCCGATCCACACAGGCCA 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-995-973-1
; Sequence 1, Application US/09995973
; Publication No. US20030024006A1
; GENERAL INFORMATION:
; APPLICANT: CHOO, Yen
; APPLICANT: ULLMAN, Christopher G.
; TITLE OF INVENTION: GENE SWITCHES
; FILE REFERENCE: 8325-2003 / G7-051
; CURRENT APPLICATION NUMBER: US/09/995.973
; CURRENT FILING DATE: 2002-03-19
```

```
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 995
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: TFIITIA/Zif-VP16
US-09-995-973-1

Query Match          12.6%; Score 152.4; DB 9; Length 995;
Best Local Similarity 57.6%; Pred. No. 1,9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 208 GCCCGCCCGCGCTCCCGAGAGGTCATCTGCTCTCCCTGACGCGCCCAATTA 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26  GCGCGTGGCGGTGGTATAGCGGTACATCTGCTCTTCCCGACGCGCGCTTA 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 268 CAGCAAGCCCTGGAACTGTGACGGGACCTGTGCAAGCAGCGGGGAGAACCATTTGT 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86  TAAACAAGAACTGAAACTGCAAGCGCATCTGTGCAAAACACAGAGAGAAACCATTTCC 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 328 TTGTGACTATGAAGGTTGTGCGAGCGCTTCATCAGGAGTACCATGTCGAGCCGACAT 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146  ATGTAGAGAGAGAGATGTGAGAAAGCTTTACCTCGCTTATACCTTAACCGGCACATC 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 388 TCTGACTCACACAGAGAGAAAGCGTTTGTGTGACAGCCACTGGCTGTGATCAAAAT 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206  ACTGACTCATCTGCGCGAGAAACCTTCACATGTGACGTGGATGATGTGATGAGATT 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 448 CAACACAATCAAACTTGAAGAAACATTTTGAAGCAACATGAATCAACAAAACA 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266  TACTACAAAGGCAACATGAAGACACTTTAAGAGATTTCATTAACATCAAGATCTGCGT 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 508 ATATATATGAGTTTGAAGACTGTAAAGACCTTTAAGAAACATGACAGCTGAAAT 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326  CTATGTGTGCCATTTTGAAGACTGTGCAAGCATTCAGAGAACATCAATTAAGGT 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 568 CCATCAGTCCAGCATACCAATGAACCTTATTCAGTGTACCCAGAGAGATGTGGAA 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386  TCATCAGTTCAGTCAACACAGACAGCTGCCGTATGCTTGCCTGTGCGATCG 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 628 ACACATTGATCATCCAGCAAGCTGAAAGCATGCCAAGGCCCAACAGGGCTA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446  CCGGTTTTCGTGCTGGATGAGCTTACCCGCCATATCCGATCCACACAGGCCA 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-732-348-4
; Sequence 4, Application US/09732348
; Patent No. US20020046419A1
; GENERAL INFORMATION:
; APPLICANT: Yen Choo, et al.
; TITLE OF INVENTION: Regulated Gene Expression in Plants
; FILE REFERENCE: 674538-2001
; CURRENT APPLICATION NUMBER: US/09/732.348
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 995
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (15)..(117)
; OTHER INFORMATION: translational initiating ATG
; NAME/KEY: misc_feature
; LOCATION: (16)..(416)
; OTHER INFORMATION: Fingers 1 to 4 of TFIITIA
; NAME/KEY: misc_feature
; LOCATION: (308)..(416)
```

```
OTHER INFORMATION: spacer
NAME/KEY: misc.feature
LOCATION: (417)..(689)
OTHER INFORMATION: three fingers of zinc fingers protein zif268
NAME/KEY: misc.feature
LOCATION: (701)..(722)
OTHER INFORMATION: Nuclear Localization Signal
NAME/KEY: misc.feature
LOCATION: (957)..(986)
OTHER INFORMATION: c-myc tag
US-09-732-348-4
```

```
Query Match 12.6%; Score 152.4; DB 10; Length 995;
Best Local Similarity 57.6%; Pred. No. 1.9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
```

```
QY 208 GCGCGCGCGCGCGCTCCGAGAGGTTGCTGCTCCCTCCGCTGACGAGCCCAATTA 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26 GCGCGCTCCGCGTGTATAGCGCTACATCTGCTTTGCCGACTGCGCGCTGCTTA 85
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 CAGCAAGCGTGGAGCTTGACGCGACCTGTGCAAGCAGCAGGGGAGAGACCTTGT 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 TACCAAACTGAAACTGCAAGCGGCGCTGTGCAAAACACAGAGAGAAACCATTTCC 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 TTGTGACTATCAAGCGTGTGCAAGCGCTTCATCAGGAGTACCATCTGAGCCGCACAT 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 ATGTAAAGAAAGAGATGTGAAGAGCTTTACCTGCTTCATCACTTAACCGGCATC 205
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 TGTGACTCACACAGAGAAAGCGCTTTGTTGTGACGCCACGCTGCTGTCAAAATT 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 ACTGACTCACTGCGCAAGAAACCTTCACATGTGACTCGAGATGCTGTGAGATT 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 448 CAACACAAATCAAACTTGAAGAAACATTTTGAACGCAACATGAATCAACAAACA 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266 TACTACAAAGCAAACTGAAGAAAGCACTTTACAGATTTCATACATCAAGATCTCGT 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 508 ATATATATGAGTTTGAAGACTGTGAAGACCTTTAAGAAACATGACGAGTGAAT 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326 CTATGTGTGCAATTTGAGAACTGTGCAAGCAATCAAGAAACATCAATTAATTAAGT 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 568 CCATGACTGCCAGCATTCGCAATGAACCTTATTCATGATGACCAAGAGATGTGGAA 627
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386 TCATCACTTGTAGTACACACAGCAGCTGCGTATGCTTGCCTGTCAGATCTCGCATCG 445
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 628 ACATTTGATCAACCGCAAGCTGAAGAGACATGCCAAGCGCCAGAGGCTA 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 CCGCTTTTCGCTCGATGAGCTTACCGCATATCCGATCCACACAGGCCA 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 13

```
US-09-783-590-5443
Sequence 5443, Application US/09783590
Patent No. US2002010850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OR INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: Po-16.2c1
CURRENT APPLICATION NUMBER: US/09/783.590
PRIOR APPLICATION NUMBER: 2000-02-15
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 5443
LENGTH: 449
TYPE: DNA
ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: misc.feature
LOCATION: (8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (53)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (66)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (83)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (181)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (245)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (312)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (359)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (364)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (397)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (448)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5443
```

```
Query Match 11.7%; Score 141.6; DB 10; Length 449;
Best Local Similarity 79.9%; Pred. No. 1e-28;
Matches 298; Conservative 0; Mismatches 58; Indels 17; Gaps 11;
```

```
QY 418 TTGTGACGACCTGGC-TGTGATCAAAATTCACACAAATCAAACTTGAGAAACATT 476
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 TTGNGCAGCCAAATGCTTGANATCAAAATTCACACAAATCAAACTTGAGAAACATT 122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 477 TTG-TACGCAAACTGAATAATCAAAACAAATATATATG-CAGTTTGAAGCTGTAA 534
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 TTGAAAGCAAACTGAATAATCAAAACAAATATATATGCTTTTGAAGCTGTAA 182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 535 -GAAGACCTTTAAGAAA-----CATCAGCAGCTGAATAATCCATCATGCGCATACCAA 588
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 GGAAGACCTTTAAGAAACATTCAGAGCTGGAAATTCATCCAGTGGCCAGCATACCAA 242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 TG-AACCTGATT-CAAGGTACCA-GGAAGATGTGGGAACACTTTG--CATCACCC 643
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 TGNAACTCTATTCACAGTATCCAGTGAAGGATGTGGGAACACTTTGCAATTCACCC 302
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 644 AGCAAGCTGAAA--CGACATGCCAAGGCCACGAGGCTATGTTG-TCAAAAAAGATGT 700
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 AGCAAGCTGNAAGACGACATGCGAAGGCCACGAGGCTTATGATGTTCAAAAAAGATGT 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 701 TCTTTTGGCAAAACATGAGCAGCACTTCTGAACATGTGAGAGAAACCATTAAGAG 760
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 TNCNTTGTGGCAAAACATGAGGAGGACCTTCTGNACTGTGAGNGAACCTTAAGGG 422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 761 GAAATTAATGTG 773
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 GATTATCTGTGAG 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
US-09-833-381-1175
; Sequence 1175, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIORITY FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1175
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1111)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1175
```

```
Query Match          7.1%; Score 86.6; DB 10; Length 1111;
Best Local Similarity 61.1%; Pred. No. 1.3e-13;
Matches 140; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
```

```
QY 186 GCGAGAGCTCAGCTCGACCCGCCGCCGCTTCCCGAGAGGTTTCATCTGCTCCT 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4 GCGTCCGCCGCCGCCGCCGCCGCCCTCGCGCCGCCCGCCAGAGCCACTGCTGCTCCT 63

QY 246 TCCTGACTGAGCGCCCAATTACAGCAAAAGCCTTGAAAGCTTGACGCCACCTGTGCAAGC 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 TCCCGGACTGGCCCAAGCCCTACTCAAGTCTCGACCTAAAGTCCGACCTCGCGAGCGC 123

QY 306 ACAGCGGGGAGAGACCATTTGTTGTGACTGTGAAGGTTGCGCAAGGCTTCATCAAGGG 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124 ACACAGGGGAGAGCGCCCTTTGCTGTGTGACTGGCAGGGCTGCGACAAAGATTGCCCGCCT 183

QY 366 ACTACACTGTGAGCGCCGACATTCGACTACACAGAGAGAAAGCCGTT 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 CCGAGCAGCTGCGCCGCCGACCGAGCGAGCGGAGAGAGCGCTT 232
```

```
RESULT 15
US-09-920-300A-1747/C
; Sequence 1747, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; PRIORITY FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1747
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 223..323
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-1747
```

```
Query Match          6.7%; Score 81; DB 10; Length 368;
Best Local Similarity 98.8%; Pred. No. 2.2e-12;
Matches 81; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 677 GCGTATGTATGTCAAAAAGATGTTCTTGTGGCAAAAACATGAGAGAACTTGTGAAA 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 GCGTATGTATGTCAAAAAGATGTTCTTGTGCAAAAACATGAGAGAACTTGTGAAA 297

QY 737 CATGTGAGAGAAACCATTAAG 758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 296 CATGTGAGAGAAACCATTAAG 275
```

```
Search completed: February 10, 2003, 12:58:42
Job time : 74.895 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:48:16 : Search time 3221.82 Seconds
(without alignments)
10957.076 Million cell updates/sec

Title: US-09-831-426c-4

Perfect score: 1213

Sequence: 1 gfgccgcgcgcgcgcgaag.....cagtactacccttgctaa 1213

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pin:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgt_hum:*
40: em_hgt_mus:*
41: em_hgt_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1171.6	96.6	1381	9	HUMH11A	D32257 Human GTF3A
2	1170	96.5	1399	6	E10962	E10962 CDNA's enco
3	1088.8	89.8	1173	9	HSU20272	U20272 Human DNA/R
4	919.2	75.8	1098	9	HSU14134	U14134 Human trans
5	889.8	73.4	1920	9	AK057993	AK057993 Homo sapi
6	719.2	59.3	1298	10	AF391799	AF391799 Mus muscu
7	691	57.0	1071	10	AF391798	AF391798 Rattus no
8	560.6	46.2	996	10	BC032292	BC032292 Mus muscu
9	470.2	38.8	97979	9	AC004739	AC004739 Homo sapi
10	470.2	38.8	193126	9	AC006045	AC006045 Homo sapi
11	452.6	37.3	111722	9	AC074390	AC074390 Homo sapi
12	452.6	37.3	206622	2	AC016444	AC016444 Homo sapi
13	447	36.9	199517	9	AL353741	AL353741 Human DNA
14	418.4	34.5	439	6	AX396111	AX396111 Sequence
15	416.8	34.4	439	6	AX396474	AX396474 Sequence
16	385.4	31.8	163421	2	AC012664	AC012664 Homo sapi
17	385.4	31.8	197709	9	AC073641	AC073641 Homo sapi
18	373.4	30.8	200885	2	AC012460	AC012460 Homo sapi
19	372.4	30.7	1693	9	AF265440	AF265440 Homo sapi
20	361.4	29.8	1518	5	XELFF111A	K02938 X.laevis 5S
21	355	29.3	1377	5	XBRF3A	X17695 Xenopus bot
22	350.8	28.9	1331	5	XELFINAB	M32472 X.borealis
23	340.6	28.1	1297	5	RPTRE11A	X58369 R.pilpiens m
24	340.6	28.1	1305	5	RANPRT11A	M85211 Rana pipien
25	319.2	26.3	123905	2	AC116311	AC116311 Homo sapi
26	319.2	26.3	157165	2	AC104112	AC104112 Homo sapi
27	319.2	26.3	171007	2	AC021381	AC021381 Homo sapi
28	319.2	26.3	171705	2	AC113377	AC113377 Homo sapi
29	314.4	25.9	166141	9	AL137059	AL137059 Human DNA
30	284.6	23.5	1314	5	BATPFI1A	X58366 B.americanu
31	278.6	23.0	67190	2	AC117566	AC117566 Mus muscu
32	234	19.3	234	6	AX396395	AX396395 Sequence
33	221.6	18.3	1129	5	IPOOTF3A	Z68499 I.punctatus
34	202	16.7	162774	2	AC121381	AC121381 Rattus no
35	202	16.7	182326	2	AC125693	AC125693 Rattus no
36	197.6	16.3	112902	9	AC005230	AC005230 Homo sapi
37	177	14.6	272	6	AX322436	AX322436 Sequence
38	163	13.4	167342	2	AC124828	AC124828 Mus muscu
39	163	13.4	217225	10	AL513345	AL513345 Mouse DNA
40	152.4	12.6	660	6	AX202580	AX202580 Sequence
41	152.4	12.6	947	6	AX053174	AX053174 Sequence
42	152.4	12.6	947	6	AX059770	AX059770 Sequence
43	152.4	12.6	947	6	AX201763	AX201763 Sequence
44	152.4	12.6	947	6	AX201936	AX201936 Sequence
45	152.4	12.6	995	6	AX053173	AX053173 Sequence

ALIGNMENTS

RESULT 1
LOCUS HUMH11A
DEFINITION Human GTF3A mRNA for Xenopus transcription factor IIIA homologue,
complete cds.
ACCESSION D32257
VERSION D32257.1 GI:1000446
KEYWORDS GTF3A; Xenopus transcription factor IIIA homologue.
SOURCE Homo sapiens cDNA to mRNA, clone_11b:11b library of T.Fujiwara, S.Shin
and Y.Nakamura clone:39H11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1381)
AUTHORS Arakawa,H., Nagase,H., Hayashi,N., Ogawa,M., Nagata,M.,

PF	05-SEP-1994 JP 19994211022	
PI	FUJIMARA TSUTOMU, TAKEDA SEI, SHIMADA YOSHIKAZU, OZAKI KOICHI,	
PI	SHIN TEIKIN	
PC	C12N15/09//A61K31/770.A61K48/00.C07H21/04;	
CC	strandedness: Single;	
CC	topology: Linear;	
EH	key	Location/Qualifiers
EH		
FT	source	1..1399
FT		/organism='Homo sapiens'
FT	5'UTR	1..19
FT	CDS	20..1288
FT		/product='hTFIIIA protein'
FT		1289..1399.
FT	3'UTR	location/Qualifiers
		1..1399
		/organism='Homo sapiens'
		/db_xref='taxon:9606'
BASE COUNT	405 a	349 c 354 g 291 t
ORIGIN		

Query Match	96.5%;	Score 1170;	DB 6;	Length 1399;
Best Local Similarity	99.3%;	Pred. No. 1,1e+252;		
Matches 1206;	Conservative	0;	Mismatches 5;	Indels 3; Gaps 3

[illegible]

Qy	721	GACGCAACTTTCTGAAACATGTGAGAGGAACCATTAAAGAGAAATACATATGTGAAGTATG	780
Db	799	GACGCAACTTCTGAAACATGTGAGAGGAACCATTAAAGAGAAATACATATGTGAAGTATG	858
Qy	781	CCGGAAAAACATTTAAAGCGAAGATTACCTTAAAGACACATGAAAACTCATGCCCCAGA	840
Db	859	CCGGAAAAACATTTAAAGCGAAGATTACCTTAAAGACACATGAAAACTCATGCCCCAGA	918
Qy	841	AAGGATATATGTGCTGTGCCAAGAGAAAGCGTGTGAGAAACCTTACTAGTGTTTAA	900
Db	919	AAGGATATATGTGCTGTGCCAAGAGAAAGCGTGTGAGAAACCTTACTAGTGTTTAA	978
Qy	901	TCCTCCAAAGCCATATCTCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGAACATGC	960
Db	979	TCTCCAAAGCCATATCTCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGAACATGC	1038
Qy	961	TGGCGTGGCAAAAACATTTGGCAATGAACAAAGTCACTAGGCAATGGCTGTGTGCATGA	1020
Db	1039	TGGCGTGGCAAAAACATTTGGCAATGAACAAAGTCACTAGGCAATGGCTGTGTGCATGA	1098
Qy	1021	TCCTGACAGAAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGAAAAAC -GCACTTTGG	1079
Db	1099	TCCTGACAGAAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGAAAAACGGAATTTGG	1158
Qy	1080	CCTCTCATCTCAGTGGATATATCCCTCCCAAAAGAAACAAGGGCAAGGCTATCTTTGT	1139
Db	1159	CCTCTCATCTCAGTGGATATAT -CCTCCCAAAAGAAACAAGGGCAAGGCTATCTTTGT	1217
Qy	1140	GTCAAAAGGAGAGTCAACCCAACTGTGTGGAAGACAAAGTGTCTCGACAGTTGCAGTAC	1199
Db	1218	GTCAAAAGGAGAGTCAACCCAACTGTGTGGAAGACAAAGTGTCTCGACAGTTGCAGTAC	1277
Qy	1200	TTACCTTGGCTAA 1213	
Db	1278	TTACCTTGGCTAA 1291	

LOCUS	LOCUS	LOCUS	LOCUS
HSU20272	1173 bp	mRNA	linear
Human DNA/RNA-binding protein mRNA, partial cds.			
DEFINITION	U20272		
VERSION	U20272.1	GI:644870	
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
1 (sites)			
Diew,P.D., Nagle,J.W., Canning,R.D., Ozato,K., Biddison,W.E. and			
Becker,K.G.			
Cloning and expression analysis of a human cDNA homologous to			
Xenopus TFIIA			
Gene 159 (2), 215-218 (1995)			
JOURNAL			
MEDLINE			
PUBMED			
2 (bases 1 to 1173)			
Becker,K.G.			
Direct Submission			
Submitted (25-JAN-1995) Kevin G. Becker, Nat. Inst. of Neur.			
Diseases and Stroke/NIH, Neuroimmunology Branch, 9000 Rockville			
Pike, Bethesda, MD 20892, USA			
location/Qualifiers			
1..1173			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="CH2-34.10"			
/sex="female"			
/tissue-type="brain; hippocampus"			
/clone_lib="Stratagene #936205"			
/dev_stage="2 years old"			
<1..1092			
/note="similar to Xenopus TFIIA encoded by SwissProt			

Accession Number P03001"
/codon_start=1
/product="DNA/RNA-binding protein"
/protein_id="AA75623.1"
/db_xref="GI:644871"
translation="PPAVVAESVSLTIADAFIAGESSAPPPPALPDRFICSPFD
CSANYSKARKLDALHLCHTGERFPCVYGCCGCAPIRDYHLSHILITHGEKPFVCA
NSCDQFNTKSNLKHKEKRNENQKQYICSPEDCKTKRKHOOLKIHOCCHNEPLF
KCTQEGCGHFAFPSKLRKRAHNEGYVQKCGSPVAKTMTLLKIVREHKEELICE
VCRKTRKRDYLKQHKMTAPERNQCRGCGRTYTFNLSHISLSPHESRPFV
CEHAGCKTFAMKQSLTRIAVYHDPDKKKMKLKVKSREKRSLSHLSGIIPKRRQG
QGLSLQNGESPVEDKMLSTVAVLTLG"
misc_feature
118..184
/note="encodes zinc-finger"
misc_feature
208..274
/note="encodes zinc-finger"
misc_feature
298..367
/note="encodes zinc-finger"
misc_feature
394..460
/note="encodes zinc-finger"
misc_feature
484..550
/note="encodes zinc-finger"
misc_feature
565..631
/note="encodes zinc-finger"
misc_feature
649..709
/note="encodes zinc-finger"
misc_feature
736..805
/note="encodes zinc-finger"
misc_feature
829..895
/note="encodes zinc-finger"
misc_feature
829..895
/note="encodes zinc-finger"
BASE COUNT 363 a 283 c 265 g 262 t
ORIGIN

Query Match 89.8% Score 1088.8; DB 9; Length 1173;
Best Local Similarity 99.8%; Pred. No. 1.9e-234;
Matches 1090; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 122 CCGCGCGCGTGTGCGCGAGTCGTCGTCCTTGCATCGCGCAGCGCTTCATTGCA 181
|||
Db 1 CCGCGCGCGTGTGCGCGAGTCGTCGTCCTTGCATCGCGCAGCGCTTCATTGCA 60
QY 182 GCGCGCGAGAGCTAGCTCCGACCCCGCGCGCGCGCTTCCGAGAGTTCATCTGC 241
|||
Db 61 GCGCGCGAGAGCTAGCTCCGACCCCGCGCGCGCGCTTCCGAGAGTTCATCTGC 120
QY 242 TCCCTCCCTACAGTCAGCGCAATTAAGAAAGCTTGACGGCAGCTCTGC 301
|||
Db 121 TCCCTCCCTACAGTCAGCGCAATTAAGAAAGCTTGACGGCAGCTCTGC 180
QY 302 AACGACAGCGGAGAGACCATTTGTTGTGACTATGAGGCTGTGCAAGCCCTTCATC 361
|||
Db 181 AACGACAGCGGAGAGACCATTTGTTGTGACTATGAGGCTGTGCAAGCCCTTCATC 240
QY 362 AGGAGCTACCATGTGACCGCCACATTTCTGACTACACAGAGAAAAGCCGTTTGTGT 421
|||
Db 241 AGGAGCTACCATGTGACCGCCACATTTCTGACTACACAGAGAAAAGCCGTTTGTGT 300
QY 422 GCAGCCCTGGCTGTGATCAAAAATTCACACAAAATTCGAAAGAAATTCGAAATTCGAA 481
|||
Db 301 GCAGCCCTGGCTGTGATCAAAAATTCACACAAAATTCGAAAGAAATTCGAAATTCGAA 360
QY 482 CGCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAA 541
|||
Db 361 CGCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAA 420
QY 542 TTTTAGAACAATCAAGAGTGAATTCAGTGCAGAGATACCAATGAACCTCTATTC 601
|||
Db 421 TTTTAGAACAATCAAGAGTGAATTCAGTGCAGAGATACCAATGAACCTCTATTC 480
QY 602 AAGGTACCCAGAAAGATGTGGAAACACTTTTGATCACCAGCAAGCTGAAAGCAGAT 661
|||
Db 481 AAGGTACCCAGAAAGATGTGGAAACACTTTTGATCACCAGCAAGCTGAAAGCAGAT 540

QY 662 GCCAAGCGCCACGAGGCGTATGTATGTCAAAAAGATGTCCTTGTGGCAAAAACATGC 721
|||
Db 541 GCCAAGCGCCACGAGGCGTATGTATGTCAAAAAGATGTCCTTGTGGCAAAAACATGC 600
QY 722 ACCGAATCTTGAACATGTGAGAGAAACCATTAAGAGAAATACATGTGAATGTC 781
|||
Db 601 ACCGAATCTTGAACATGTGAGAGAAACCATTAAGAGAAATACATGTGAATGTC 660
QY 782 CGGAAACATTTAAAGCAAAAGATTAACCTTAAGCAACATGAAACATCTAGTCCCGCA 841
|||
Db 661 CGGAAACATTTAAAGCAAAAGATTAACCTTAAGCAACATGAAACATCTAGTCCCGCA 720
QY 842 AGGAGTATGTGCTGTCCAGAGAAAGCTGTGGAAGAACTATACATGCTTAAAT 901
|||
Db 721 AGGAGTATGTGCTGTCCAGAGAAAGCTGTGGAAGAACTATACATGCTTAAAT 780
QY 902 CTCGAAAGCATATCTCTCTCTTCATGAGAGAAAGCCGCTTTGTGTGAACATGCT 961
|||
Db 781 CTCGAAAGCATATCTCTCTCTTCATGAGAGAAAGCCGCTTTGTGTGAACATGCT 840
QY 962 GCGTGTGGCAAAACATTTGCAATGAACAACAAAGTCTCACTAGGCATGCTGTGTATCATGAT 1021
|||
Db 841 GCGTGTGGCAAAACATTTGCAATGAACAACAAAGTCTCACTAGGCATGCTGTGTATCATGAT 900
QY 1022 CTTGACAAGAAAGAAATGAAGCTCAAGTCAAAAATCTGTGCAAAAACGAGTGTGGCC 1081
|||
Db 901 CTTGACAAGAAAGAAATGAAGCTCAAGTCAAAAATCTGTGCAAAAACGAGTGTGGCC 960
QY 1082 TCTCATCTCAGTGAATATTCCTCCCAAAAAGAAAGAGGAGGCAAGCTTCTTTGTGT 1141
|||
Db 961 TCTCATCTCAGTGAATATTCCTCCCAAAAAGAAAGAGGAGGCAAGCTTCTTTGTGT 1020
QY 1142 CAAGAAGGAGATCAACCACTGTGTGGAAGACAAGATGCTTCGACAGTTCAGTACTT 1201
|||
Db 1021 CAAGAAGGAGATCAACCACTGTGTGGAAGACAAGATGCTTCGACAGTTCAGTACTT 1080
QY 1202 ACCCTTGCTTAA 1213
|||
Db 1081 ACCCTTGCTTAA 1092

RESULT 4
HSU14134 1098 bp mRNA linear PRI 01-OCT-1994
LOCUS
DEFINITION Human transcription factor IIA (HFIIA) mRNA, partial cds.
ACCESSION U14134
VERSION U14134.1 GI:551534
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1098)
AUTHORS Drew,P.D., Nagle,D.W., Canning,R.D., Ozato,K., Biddison,W.E. and
Becker,K.G.
TITLE Cloning and expression analysis of a cDNA encoding human TFIIA
JOURNAL unpublished
REFERENCE
2 (bases 1 to 1098)
AUTHORS Becker,K.G.
TITLE Direct Submision
JOURNAL Submitted (29-AUG-1994) Kevin G. Becker, National Institute of
Neurological Diseases and Stroke/NIH, Neuroimmunology Branch, 9000
Rockville Pike, Bethesda, MD 20892, USA
FEATURES
source
1..1098
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CZH2-34"
/sex="2 yr old female"
/tissue_type="brain"
/clone_lib="Human hippocampus cDNA library, Stratagene
Number 936205"
1..1098

gene

CDS

/gene="HrFlIIA"
<1..1017
/gene="HrFlIIA"
/note="HrFlIIA: similar to Xenopus transcription factor
IIIA, Swiss-Prot Accession Number P03001"
/codon_start=1
/product="transcription factor IIIA"
/protein_id="AA21873.1"
/db_xref="GI:551535"
/translation="pPAVVAVESVSLTIADAFIAGSSAPTPRPALPRPFISSPD
CSANYSKAMKLDHLCKHTGERPVCDEECGKAFIDYHLSHILHTHEKFPVCA
NGCDQFNYSNLKHEERKHEHQKOYIGSFEDCKRPFKHOOLKHOCQHNPELE
KLCCOKGCSFVAKTWELKHKRETHKEILCEVCKRTEKRDYLOKHMHTAPEEDY
CRPRREGCRITTYTENVLOSILISFHEESRPVCEHAGCKTFAKMSLIRHAYVHD
DKKMKLVKKSREKSLASHLSGIIPKRRKQOGLSLCQNGSPNCEVDKMLSTAV
LTLC"

misc_feature
118..186
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
208..276
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
298..369
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
394..462
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
490..558
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
574..636
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
661..732
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
754..822
/gene="HrFlIIA"
/note="encodes zinc-finger"

misc_feature
/note="encodes zinc-finger"

BASE COUNT 339 a 263 c 244 g 252 t
ORIGIN

Query Match 75.8%; Score 919.2; DB 9; Length 1098;
Best Local Similarity 92.4%; Pred. No. 2.8e-196;
Matches 1009; Conservative 0; Mismatches 8; Indels 75; Gaps 1:

QY 122 CCGCGGCGCTGTGTCGCGAGTCGTGTCCTTGACCATGCGCGCTTGATGCA 181
D 1 CCGCGGCGCTGTGTCGCGAGTCGTGTCCTTGACCATGCGCGCTTGATGCA 60

QY 182 GCCGCGAGAGCTACGTCGCAACCCCGCGCGCGCTTCCAGAGGTTCAATGTC 241
D 61 GCCGCGAGAGCTACGTCGCAACCCCGCGCGCGCTTCCAGAGGTTCAATGTC 120

QY 242 TCCTTCCCTGACTGAGCGCAATTAAGCAAGGCTTGAGCGCACCTGTCG 301
D 121 TCCTTCCCTGACTGAGCGCAATTAAGCAAGGCTTGAGCGCACCTGTCG 180

QY 302 AAGCACAGGGGAGAGACATTTGTTGTGACTATGAAGGCTTGCGAAGCCCTTCATC 361
D 181 AAGCACAGGGGAGAGACATTTGTTGTGACTATGAAGGCTTGCGAAGCCCTTCATC 240

QY 362 AGGCACTACATCTGAGCGCGCAATTTGACTACAGAGAGAAAGCCCTTTGTTGT 421
D 241 AGGCACTACATCTGAGCGCGCAATTTGACTACAGAGAGAAAGCCCTTTGTTGT 300

QY 422 GCAGCCACTGCTGTGATCAAAATTAACAACAAATCAAACTGAAGAATCAATTTGAA 481
D 301 GCAGCCACTGCTGTGATCAAAATTAACAACAAATCAAACTGAAGAATCAATTTGAA 360

QY 482 CGCAAAACATGAATAACAAACAAATATATATGACATTTTGAAGACTGTAAGAAGAC 541
D 361 CGCAAAACATGAATAACAAACAAATATATATGACATTTTGAAGACTGTAAGAAGAC 420

D 361 CGCAAAACATGAATAACAAACAAATATATATGACATTTTGAAGACTGTAAGAAGAC 420

QY 542 TTATAAGAAACATGACAGCGTGAATAATCCATCAGTCCAGCATACCATGTAACCTTATTC 601
D 421 TTATAAGAAACATGACAGCGTGAATAATCCATCAGTCCAGCATACCATGTAACCTTATTC 480

QY 602 AAGTGTACCCGAGGAGGATGTGGGAAACACTTTGATCACCACAGCAAGCTGAAGACAT 661
D 481 AAG-----CTATGTTGCAAAAAAGAGATGTTCTTTGGCAAAAAACATGCG 525

QY 662 GCCAAGGCCACAGAGCGCTATGATGTCAAAAAGAGATGTTCTTTGGCAAAAAACATGCG 721
D 484 -----CTATGTTGCAAAAAAGAGATGTTCTTTGGCAAAAAACATGCG 525

QY 722 ACGGACTCTGAAACATGTGAGAGAAACCATTAAGAGAAATACATATGTAAGTATGCG 781
D 526 ACGGACTCTGAAACATGTGAGAGAAACCATTAAGAGAAATACATATGTAAGTATGCG 585

QY 782 CGGAAACATTTAAACCAAGATTACCTTAAGCAACATGAAACATGATGCCCGAGAA 841
D 586 CGGAAACATTTAAACCAAGATTACCTTAAGCAACATGAAACATGATGCCCGAGAA 645

QY 842 AGGATGTATGTGCGCTGTCGAAGAGAGCTGTGGAAGAACCTTATCTACTGTGTTAAT 901
D 646 AGGATGTATGTGCGCTGTCGAAGAGAGCTGTGGAAGAACCTTATCTACTGTGTTAAT 705

QY 902 CTCGAAAGCCATATTCCTCTCTCCATGAGAGAAAGCCGCTTTGTTGTGTGAACATGCT 961
D 706 CTCGAAAGCCATATTCCTCTCTCCATGAGAGAAAGCCGCTTTGTTGTGTGAACATGCT 765

QY 962 GCGTGTGCAAAACATTTGCAATGAACAAAGTCTCAGTACGATGCTGTTGTCATGAT 1021
D 766 GCGTGTGCAAAACATTTGCAATGAACAAAGTCTCAGTACGATGCTGTTGTCATGAT 825

QY 1022 CCGTGAACAAGAAATGAACCTCAAGTCAAAAAATCTGTGAAAAACGAGATTTGGCC 1081
D 826 CCGTGAACAAGAAATGAACCTCAAGTCAAAAAATCTGTGAAAAACGAGATTTGGCC 885

QY 1082 TCTCATCTCACTGATATATTCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCTTTGTGT 1141
D 886 TCTCATCTCACTGATATATTCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCTTTGTGT 945

QY 1142 CAAACGAGAGATCACCAACTGTGTGAGAGCAAGATGCTCTGACAGTTCAGTACTT 1201
D 946 CAAACGAGAGATCACCAACTGTGTGAGAGCAAGATGCTCTGACAGTTCAGTACTT 1005

QY 1202 ACCCTTGGCTAA 1213
D 1006 ACCCTTGGCTAA 1017

RESULT 5
AK057993 1920 bp mRNA linear PRI 31-OCT-2001
LOCUS AK057993
DEFINITION Homo sapiens cDNA FLJ25264 f1s, clone STW05057, highly similar to
TRANSCRIPTION FACTOR IIIA.
ACCESSION AK057993
VERSION AK057993.1 GI:16553988
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens stomach mucosa cDNA to mRNA, clone_11b:STM
clone:STM05057.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
Ono,Y., Hotuba,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
Yamashita,H., Chiba,Y., Suzuki,Y., Hata,H., Nakagawa,K., Mizuno,S.,
Morinaga,M., Kawamura,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
Nishikawa,T., Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and
Sugano,S.
NEBO human cDNA sequencing project

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1920)
AUTHORS Sugano,S. and Suzuki,Y.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction and 5'-end one pass sequencing; Institute of Medical
Science, University of Tokyo, Laboratory of Genome Structure, Human
Genome Center; 3'-end one pass sequencing; RAB; clone selection for
full insert sequencing; RAB and Helix Research Institute.
FEATURES
Source
1..1920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="STM05057"
/tissue.type="stomach mucosa"
/clone.lib="STM"
/note="cloning vector: pME18SFL3"
BASE COUNT 520 a 462 c 438 g 500 t
ORIGIN
Query Match 73.4%; Score 889.8; DB 9; Length 1920;
Best Local Similarity 99.2%; Pred. No.1.2e-189;
Matches 891; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 313 GGACAGACATTTGTTGTGACTATGAGGGGTGGCAGGCCCTTCATCAGGACTACCA 372
|||
DB 928 GGTGAGACATTTGTTGTGACTATGAGGGGTGGCAGGCCCTTCATCAGGACTACCA 987
QY 373 TCTGAGCCGCCACATTTGACTCAGACAGAGAAACCGTTTGTGTCAGGCACTGG 432
|||
DB 988 TCTGAGCCGCCACATTTGACTCAGACAGAGAAACCGTTTGTGTCAGGCAATGG 1047
QY 433 CTGTGATCAAAAATTCAGACAAAATCAAACTTGAGAAACATTTGAGCAACATGA 492
|||
DB 1048 CTGTGATCAAAAATTCAGACAAAATCAAACTTGAGAAACATTTGAGCAACATGA 1107
QY 493 AAATCAACAAAACATATATATGTCAGCTTTGAAAGACTGTAGAGACCTTTAAGAAACA 552
|||
DB 1108 AAATCAACAAAACATATATATGTCAGCTTTGAAAGACTGTAGAGACCTTTAAGAAACA 1167
QY 553 TCAGCAGCTGAAAATTCATCAGTCCGAGCATACCAATGAACCTTATTCAAGTACCCA 612
|||
DB 1168 TCAGCAGCTGAAAATTCATCAGTCCGAGCATACCAATGAACCTTATTCAAGTACCCA 1227
QY 613 GGAAGAGTGGGAAACATTTGATCAGCCAGCAGCACTGAAAGCATGGCAAGGCCCA 672
|||
DB 1228 GGAAGAGTGGGAAACATTTGATCAGCCAGCAGCACTGAAAGCATGGCAAGGCCCA 1287
QY 673 CGAGGCGTATGTATGTCAAAAGAGATGTCCTTTGTGCAAAAACATGACGGAACCTCT 732
|||
DB 1288 CGAGGCGTATGTATGTCAAAAGAGATGTCCTTTGTGCAAAAACATGACGGAACCTCT 1347
QY 733 GAAACATGTAGAGAAACCCATTAAGAGAAATCTATGTGAGTATGCCGAAACATT 792
|||
DB 1348 GAAACATGTAGAGAAACCCATTAAGAGAAATCTATGTGAGTATGCCGAAACATT 1407
QY 793 TAAAGCAAGATTACCTTAAGCAACATGAAACATCATGCCCGAAGAGGATGTATG 852
|||
DB 1408 TAAAGCAAGATTACCTTAAGCAACATGAAACATCATGCCCGAAGAGGATGTATG 1467
QY 853 TCGCTGTCAAGAGAGAGCTGTGGAAGAACCTATAGTGTGTTTAATCTCCAAAGCA 912
|||
DB 1468 TCGCTGTCAAGAGAGAGCTGTGGAAGAACCTATAGTGTGTTTAATCTCCAAAGCA 1527
QY 913 TATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGAACATCTGGCTGTGCAAA 972
|||

DB 1528 TATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGAACATCTGGCTGTGCAAA 1587
QY 973 AACATTGCAATGAACAAAGTCTCACTAGCATGCTGTTGTACATGATCCTGACAAAGA 1032
|||
DB 1588 AACATTGCAATGAACAAAGTCTCACTAGCATGCTGTTGTACATGATCCTGACAAAGA 1647
QY 1033 GAAATGAAGCTCAAAAGTCAAAAATCTCGTGAAGAAAGAGATTGGCCTCATATCTCAG 1092
|||
DB 1648 GAAATGAAGCTCAAAAGTCAAAAATCTCGTGAAGAAAGAGATTGGCCTCATATCTCAG 1707
QY 1093 TCGATATATCCCTCCCAAAAGAAACAGGCAAGCTTATCTTTGTGTAAACAGAGA 1152
|||
DB 1708 TCGATATATCCCTCCCAAAAGAAACAGGCAAGCTTATCTTTGTGTAAACAGAGA 1767
QY 1153 GTCAACCAATGCTGTGAGACAAAGATGCTCTGACAGTTCAGTACTTACCTTGCTGA 1212
|||
DB 1768 GTCAACCAATGCTGTGAGACAAAGATGCTCTGACAGTTCAGTACTTACCTTGCTGA 1827
QY 1213 A 1213
DB 1828 A 1828
RESULT 6
AF391799 1298 bp mRNA linear ROD 31-JAN-2002
LOCUS AF391799
DEFINITION Mus musculus transcription factor IIA mRNA, partial cds.
ACCESSION AF391799
VERSION AF391799.1 GI:18448381
KEYWORDS
SOURCE
Mus musculus.
ORGANISM
Mus musculus.
REFERENCE
1 (bases 1 to 1298)
AUTHORS Hanas,J.S., Hocker,J.R., Cheng,Y.G., Lerner,M.R., Brackett,D.J.,
Lighthoot,S.A., Hanas,R.J., Madhusudhan,K.T. and Moreland,R.J.,
CDNA cloning, DNA binding, and evolution of mammalian transcription
factor IIA
TITLE
JOURNAL Gene 282 (1-2), 43-52 (2002)
MEDLINE 21673987
PUBMED 11814676
REFERENCE
2 (bases 1 to 1298)
AUTHORS Hanas,J.S., Hocker,J.R., Lerner,M.R., Brackett,D.J.,
Lighthoot,S.A., Hanas,R.J., Madhusudhan,K.T. and Moreland,R.J.,
Direct Submission
TITLE
JOURNAL Submitted (14-JUN-2001) Biochemistry/Molecular Biology, University
of Oklahoma Health Science Center, 800 Research Parkway, Suite 448,
Oklahoma City, OK 73104, USA
FEATURES
Source
1..1298
/organism="Mus musculus"
/db_xref="taxon:10090"
<1..1203
/note="TFIIA: Cys2His2 zinc finger protein: similar to
Xenopus laevis transcription factor IIA Swiss-Prot
Accession Number P03001"
/codon_start=1
/product="transcription factor IIA"
/protein_id="AAL69686.1"
/db_xref="GI:18448382"
/translation="EFGTEVOSLRAPVAPRSGRHVAAROPGRDVALLEPRVSA
EAVSLLTADAFVAGCEPAPRPAPLPSPICSPCSASYNKAMKLDHILCKHGER
PFCYDEGGKATIRYHLSRHYLITGKRPFCADGDCNOKRNTSNLKHIERHG
NPQKYLCTEGCKAKFKHQDLRTHOCHTSPLRCHTEGGGKHFASPSRLRKG
VHEGYLCQKGCSPMGKTWELDKHMRBAKEDITVCQRMFKRRDYLRQHMKTAPR
RDVYRCRPGCGRTYTVFNLSHLSIFHEKRPVCEHAAGCGKTFAMKQSLRHSIV
HDDDKRMKLKVAPRRERSLASRLSGTFPPRKQKQDPVSLPMASSSSSPQAQLP
PAALLVC"
BASE COUNT 342 a 368 c 358 g 230 t
ORIGIN
Query Match 59.3%; Score 719.2; DB 10; Length 1298;

Best Local Similarity 79.5%; Pred. No. 2.8e-151;
Matches 879; Conservative 0; Mismatches 218; Indels 9; Gaps 2;

```
OY 110 GGGCCCTGGATCCGGCGCGCGTGTGCGCGAAGTCGCTGCTTGCACCATGCGCCGAC 169
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 103 GTCGCCCTGGAGGCGCGGTGTGTCAGTCGCGAAGCGGTGTGCTGCACCATGCGGAT 162

OY 170 GCGTTTCATTCAGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCTCCGAG 229
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 GCGTTTCATTCAGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCTCCGAG 216

OY 230 AGGTTTCATTCAGCTTCCCTCCCTGACGCGGCAATTACAGCAAAAGCGCTGGAAGCTTGC 289
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 AGGTTTCATTCAGCTTCCCTCCCTGACGCGGCAAGTTACAAAGCAAGCGCTGGAAGCTGAC 276

OY 290 GCGGACCTGTGCAAGCAGCAGCGGGGAGAGACCATTTGTTGTGCTATGAAGGTGTGGC 349
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 GCGGACCTGTGCAAGCAGCAGCGGGGAGAGACCATTTGTTGTGCTATGAAGGTGTGGC 336

OY 350 AAGGCTTCATCAGGAGCTACATCTGAGCGCGCCACATTCGACTCAGACAGAGAAAG 409
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 337 AAGGCTTCATCAGAGACTACATCTGAGCGCGCATGCTGATTTCACCGGGGAAAG 396

OY 410 CCGTTTGTGTGCGAGCGCAGCTGCTGTATCAAAATTCACACAAATCAAACTTGAAG 469
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 397 CCGTTTGTGTGCGAGCTGCTGTATCAAAATTCACACAAATCAAACTTGAAG 456

OY 470 AAGATTTTGAAGCAGCAACATGAAATTCACAAACAAATATATGATGATTTGAAGAC 529
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 457 AAGATTTTGAAGCAGCAACATGAAATTCACAAACAAATATATGATGATTTGAAGAC 516

OY 530 TGTAGAAGACCTTTAAGAAACATCAGAGCTGAAATTCATGCTGACCATATACCAAT 589
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 517 TGTAGAAGACCTTTAAGAAACATCAGAGCTGAAATTCATGCTGACCATATACCAAT 576

OY 590 GAACCTTATTAAGGTGTACCCAGAAAGATGTGGAAACATTTGATCACCAGCAAG 649
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 577 GAGCGCTCTTCAGGTGTACCCAGAAAGATGTGGAAACATTTGATCACCAGCAAG 636

OY 650 CTGAAAGACATGCGCAAGGCCAGAGAGGCTATGATGTCAAAAAGATGTTCTTGTG 709
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 637 CTGAAAGACATGCGCAAGGCCAGAGAGGCTATGATGTCAAAAAGATGTTCTTGTG 696

OY 710 GCAAAACATGAGAGCAACTTCTGAACATGTGAGAGAAACCATTAAGAGAAATACTA 769
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 697 GCAAAACATGAGAGCAACTTCTGAACATGTGAGAGAAACCATTAAGAGAAATACTA 756

OY 770 TGTGAAGTATGCGGAAACATTTAAAGCAAGATTAACCTTAAGCAACACATGAAACT 829
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 757 TGTGAAGTATGCGGAAACATTTAAAGCAAGATTAACCTTAAGCAACACATGAAACT 816

OY 830 CATGCCCCAGAAAGAGATGTGTGCGCTGTCAAGAGAGGCTGTGAAGAACTATACT 889
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 817 CATGCCCCAGAAAGAGATGTGTGCGCTGTCAAGAGAGGCTGTGAAGAACTATACT 876

OY 890 ACTGCTTTAATCTCAAAAGCATATCTCTCTTCCATGAGAAAGCGCGCTTTGTG 949
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 877 ACTGCTTTAATCTCAAAAGCATATCTCTCTTCCATGAGAAAGCGCGCTTTGTG 936

OY 950 TGTGAACATGCTGCTGTGCAAAACATTTGCAATGAAACAAAGCTTCATAGGCAATGCT 1009
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 937 TGTGAACATGCTGCTGTGCAAAACATTTGCAATGAAACAAAGCTTCATAGGCAATGCT 996

OY 1010 GTTGTATATGATCTCTGACAAAGAAATGAAGTCAAAAGTCAAAATCTCTGTGAAAAA 1069
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 997 GTTGTATATGATCTCTGACAAAGAAATGAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAG 1056

OY 1070 CGAGTTTGGCTCTCATCTCAGTGAATATATCTCCCAAAAGAAAGCAAGGCGCAAGGC 1129
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1057 CGAGTTTGGCTCTCATCTCAGTGAATATATCTCCCAAAAGAAAGCAAGGCGCAAGGC 1116

OY 1130 TTATCTTTGTGTCA --AAACGAGAGTCAACCAACTGTGTGAAGACAGATGCTCTCG 1186
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 1117 TACTCTTGGCTTAAGCGCAGCGCAGAGCTCCAGCAGACGCCAGAGGCCAGTCCCGG 1176
OY 1187 ACAGTGCAGTACTTACCTTTGGCTA 1212
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1177 CCAGCGCCTTACTACTGTCTGCTA 1202

RESULT 7
AF391798
LOCUS
DEFINITION
Rattus norvegicus transcription factor IIIA mRNA, partial cds.
VERSION
AF391798.1 GI:18448379
KEYWORDS
SOURCE
Rattus norvegicus.
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1071)
Hanas,J.S., Hocker,J.R., Cheng,Y.G., Lerner,M.R., Brackett,D.J.,
Lightfoot,S.A., Hanas,R.J., Madhusudan,K.T., and Moreland,R.J.
CDNA cloning, DNA binding, and evolution of mammalian transcription
factor IIIA
JOURNAL
Gene 282 (1-2), 43-52 (2002)
MEDLINE
21673987
PUBMED
11814676
REFERENCE
2 (bases 1 to 1071)
Hanas,J.S., Hocker,J.R., Lerner,M.R., Brackett,D.J.,
Lightfoot,S.A., Hanas,R.J., Madhusudan,K., and Moreland,R.J.
Direct Submission
Submitted (14-JUN-2001) Biochemistry/Molecular Biology, University
of Oklahoma Health Science Center, 800 Research Parkway, Suite 448,
Oklahoma City, OK 73104, USA
FEATURES
source
1..1071
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
<..1011
/note="TFIIIA: Cys2His2 zinc finger protein; similar to
Xenopus laevis transcription factor IIIA Swiss-Prot
Accession Number P03001"
CDS
1..1011
/codon_start=1
/product="transcription factor IIIA"
/protein_id="AAL69685.1"
/db_xref="GI:18448380"
/translation="AARAGALPSRFPCSPDASASYNKAKLDAHLCKHNGERSFYCD
YEGCGKAFRTDYLSRIILIHGCEKPYCADNCCNCKRSTKSLKHKHIEKHNPKO
YVCFEGCKKAEKKHOOALRTHOCQHTNEPLFRCTHSCGKHNPASPSRLKHGKVDYR
LCQKGSFVGKRTVELIKHTREAKRELEVTVQKMKRRDKLKHMKTHAPERDVYR
CPREGCARITYTFVFNLSHILSFHEKRPVCEHACGKTFPAKROSLMRSHVHDDPK
KRMLKLVPRPRERRSLASRLSGVPPKQEPDCLSPNSTRESSSSPEATWLAPELLT
VH"
BASE COUNT
294 a 283 c 290 g 204 t
ORIGIN
Query Match
57.0%; Score 691; DB 10; Length 1071;
Best Local Similarity 80.9%; Pred. No. 6.1e-145;
Matches 805; Conservative 0; Mismatches 190; Indels 0; Gaps 0;
```

```

QY 398 ACAGAGAAAAGCGTTTGTGTCAGCCACTGGCTGTATCAAAAATTCAACACAAA 457
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196 ACGGAGAAAAGCATTTTGTGTCAGATTAATGGCTGTATCAAGAAATTCAGACAAA 255
QY 458 TCAACTTGAAGAAACATTTTGAACGCAAAATCAAAAATCAATATATATG 517
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 256 TCAAACTTGAAGAAACATTTTGAACGCAAAATCAAAAATCAATATATATG 315
QY 518 AGTTTGAACACTGTGAAGACCTTAAAGAAATCAAGCAAGCAAAATCCATCATG 577
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 316 AATTTGAAGCTTGAAGAGGCTTTAAGAGCACCAGCAAGCAAGCAAGCAAGTGC 375
QY 578 CACATACCAATGAACCTTATTCATCACTTACCCAGAGAGATGTGGAAACATTTGCA 637
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 376 CAGAGACCAATGAGCAGCTTTCAGGTGTACCCAGAGAGATGTGGAAAGCACTTGCC 435
QY 638 TCACCCAGCAAGCTGAAGACATGCCAGCCAGAGGGCTATGTATGTCAAAAAGA 697
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 436 TCCCCAGCAGGCTGAAGACGAGTGGAGAGTTCACAGCGCTACCTATGTCAAAAAGGA 495
QY 698 TGTTCCTTGTGGCAAAACATGAGAGCACTTCTGAAGATGTGAAGAAACCATATA 757
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 496 TGTTCCTTGTGGCAAAACATGAGAGCACTTCTGAAGATGTGAAGAAACCATATA 555
QY 758 GAGGAAATTAATGATGAGATATGCCGGAACAAATTAAGCAAAAGATTAAGCAA 817
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 556 GAGGAGGTGACCTGACAGAGTGTGCAGAAAGATGTTCAGAGCAAAAGATCACTTAACAG 615
QY 818 CACATGAAACTCATGCCCCAGAAAGGATGTATGCTGCTCCAGAGAGAGGCTGTGA 877
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 616 CACATGAAACTCATGCCCCAGAAAGGATGTATGCTGCTCCAGAGAGGCTGTGA 675
QY 878 AGAAGCTATCTACTGTGTTAATCTCCAAAGCATATCTCTCTCCATGAGGAAAGC 937
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 676 AGAAGCTATCTACTGTGTTAATCTCCAAAGCATATCTCTCTCCATGAGGAAAGC 735
QY 938 CGCCCTTTTGTGTGAACATGCTGCTGTGCAAAACATTTGCAATGAAGCAAGTCTC 997
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 736 CGCCCAATTTGTGTGAGCATGCTGCTGTGCAAGAGAGGTTGGATGAAGAAAGTCTC 795
QY 998 ACTAAGGATCTGTGTGATGATCTGACAGAGAAATGAAGCTCAAAAGTCAAAAA 1057
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 796 ATGAGGACACGTGCTGCAAGATCTTCAGCAAGAAAGGATGAAGCTCAAAAGTGAACCA 855
QY 1058 TCTGTGAAAGAGGAGTTTGGCTCTCATCTGATGATATATCCCTCCCAAAAGAAA 1117
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 856 CCTGCGAGAGAGCAGCAGCTTGGCTCGCGCTCAGCGGTAAGTCTCTCTCAAGGGGAAA 915
QY 1118 CAAGGCGAAGGCTTATCTTGTGTCAAAACGAGAGTCAACCAAGTGTGAAGACAAG 1177
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 916 CAAGAGCGGACAGTGTCTCTGCTAAGCAGAGAGTCAAGAGCCAGAGGCGCAGC 975
QY 1178 ATGCTCTGACAGTGTGCACTTACCTTGGCTA 1212
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 976 ATGCTCTGACAGTGTGCACTTACCTTGGCTA 1010

```

```

RESULT 8
LOCUS BC032292 996 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, general transcription factor III A, clone MGC:40923
ACCESSION BC032292
VERSION BC032292.1 GI:21595507
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 996)
AUTHORS Strausberg, R.
TITLE Direct Submission

```

JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantipop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 65 Row: 1 Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source

1..996
 /location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N-3"
 /clone="MGC:40923 IMAGE:5374268"
 /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."
 /clone.lib="MCI CGAP_Mam2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 /cvs .874
 /codon_start=1
 /product="general transcription factor III A"
 /protein_id="AAH32292.1"
 /db_xref="GI:21595508"
 /translation="MGRTWTELELKHREAHKEDITCNVCFRFRDYLKOHKTHAP ERDVRCPGCGRTYTFVNLQSHLSFHEERKRPVCEHAGCGKTFAMKOSLMRHSV VHDPRKRMKLKVRAPRRRSILASRLSGVFPKRRQEPDYSIPNASASSSPDQLP PATLITVC"
 CDS

BASE COUNT 319 a 261 c 244 g 172 t

Query Match 46.2%; Score 560.6; DB 10; Length 996;
 Best Local Similarity 78.9%; Pred. No. 1.3e-115;
 Matches 681; Conservative 0; Mismatches 179; Indels 3; Gaps 1;

```

QY 353 GCCTTCATCAGGACATGACGCGCCACATTTGCTACACAGAGAAAACCG 412
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 11 GCCTTCATCAGGACATGACGCGCCACATTTGCTACACAGAGAAAACCG 70
QY 413 TTGTTGTGACGACATGCGCTGTGATCAAAAATTCAAACCAAAATCAAACTGAAGAAA 472
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 71 TTGTTGTGACGATGATGCTGTATCAAGAAATTCAAACCAAAATCAAACTGAAGAAA 130
QY 473 CATTTGAAGCAAAACATGAATAATCAACAAATATATGATGAGTTTGAAGACTGT 532
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 131 CACATTGAAGCAAAACATGGAAGAACCAAAACAGATATGTGTGCACTTATGAGGTTGC 190
QY 533 AAGAAGACCTTTAAGAAACATCAGCAGCTGAATAATCCATCATGCGCAGCATACCAATGAA 592
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 191 AAGAAGACCTTTAAGAAACATCAGCAGCTGAATAATCCATCATGCGCAGCATACCAAGCAG 250

```

OY		593	CCTCATTTCAAGTGTACCCAGGAAAGATGTGGAAACACTTTGCATTCACCAGCAGCGTG	652
Db		251	CCACTCTTAAGGTGTACCCAGGAAAGATGTGGAAACACTTTGCATTCACCAGCAGCGTG	310
OY		653	AAACGACATGGCCAAGGCCAGGAGGAGTATGTATGTCAAAAAGATGTCTCTTTGTGCA	712
Db		311	AAACGACATGGGAAGAAGTTCACAGAGGCGTACCGTGTCAAAAAGATGTCTCTTTGTGGA	370
OY		713	AAAACATGGACGGAACCTTCTGAAACATGTGAGAGAAACCCATMAAGAGAAATACTATGT	772
Db		371	AAAACGTCGACAGAGACTTCTGAAACATGTGAGAGAACCCATMAAGAGACATAACCTGC	430
OY		773	GAAATATCCCGGAAAAAATTAAAGCAAAGATTACCTTAAAGCAACATGATGAAACATAT	832
Db		431	AACGATATCTCAGAGATTTCAAGCGCAGAGATTACCTTAAAGCACACATGAAAGCTAC	490
OY		833	GCCCCAGAAAGGATGTATGTCTGCTGTCCAGAGAGAGGCTGTGAAAGAACTATACTACT	892
Db		491	GCCCGGAAAGGATGTATGTCTGCTGTCCAGAGAGAGGCTGTGAAAGAACTATACACACC	550
OY		893	GCTGTTAATCTCCAAAGCCATATCTCTCTCTTCCATGAGAAAGCCGCTTTGTGTGT	952
Db		551	GCTTCAACCTGCAAGGACACATTTCTCTCTCCAGGAGAAACCGCCATTTGTGTGT	610
OY		953	GAACATGCTGGCTGGGCAAAACATTTGCAATGAAACAAGCTCCTAGAGCATCTGTT	1012
Db		611	GAGCAGCTGGCTGGGCAAGACATTTGCCAATGAAACAGAGTCTATATAGCACAGTCTC	670
OY		1013	GTCATGATCTCTGACAGAGAAATGAAGCTCAAAAGTCAAAAAATCTGTGAAAAACGG	1072
Db		671	GTCGACGATCTCCGCAAMAAAGAGATGAAGCTCAAAAGTAAAGCCCCCTGGAGAGAGCG	730
OY		1073	AGTTGGGCTCTCATCTAGTGGATATATTCCTCCCAAAAGAAACAGGCGCAGGCTTA	1132
Db		731	AGCTTGGCTCTCGCTCAGTGGGATCTTCCCTCTTAAGAGAAACAAGACCCGACATAC	790
OY		1133	TCTTTGTGTCA---AAACGAGAGTCACCAACTGTGTGAAAGACAAAGATCTCTCGACA	1189
Db		791	TCTTGGCTTAAGGCGAGGAGAGTCAGACAGTCCAGACAGTCCAGAGGCCAGCTGC	850
OY		1190	GTTGCACTACTTACCCCTTGCTTA	1212
Db		851	GCCACCTTACTCAGTGTCTGCTA	873
RESULT 9				
LOCUS	AC004739/c	97979 bp	DNA linear	PRI 03-FEB-2000
DEFINITION	Homo sapiens PAC clone RP4-531G15 from 7p21, complete sequence.			
ACCESSION	AC004739			
VERSION	AC004739.1	GI:3152634		
KEYWORDS	HTG.			
SOURCE	Homo sapiens.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	Kemp, K.			
AUTHORS	The sequence of Homo sapiens PAC clone RP4-531G15			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 97979)			
AUTHORS	Waterston, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-MAY-1998) Department of Genetics, Washington			
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
REFERENCE	3 (bases 1 to 97979)			
AUTHORS	Waterston, R.			
JOURNAL	Direct Submission			
	Submitted (03-FEB-2000) Department of Genetics, Washington			
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
TITLE	Genome Center			
JOURNAL	-----			
COMMENT	Center: Washington University Genome Sequencing Center			

Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
Summary Statistics
Center project name: H_DJ0531615

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McNetson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPC1-4, prepared by Plierer de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 5:88-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).
JPCRCB:PCV38C3

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RP4-531G15;
actual end is at base position 97979 of RP4-531G15.

This clone contains STS SWSS2758 (NID:g113567).
Location/Qualifiers

```

/organism="Homo sapiens"
/ab_xref="taxon:3600"
/chromosome="7"
/map="7p21"
/clone="RP4-531G15"
/clone_lib="RRC1-4"
1..62
/rpt_family="Alu"
/rpt_family="L1"
137..179
/rpt_family="L2"
660..831
/rpt_family="MIR"
338..3572
/rpt_family="Alu"
3668..4067
/rpt_family="MIR"
7085..8157
/rpt_family="L1"
8397..8714
/rpt_family="Alu"
8718..9750
/rpt_family="L1"
9751..12094
/rpt_family="L1"
12095..12372
/rpt_family="Alu"
12373..12436
/rpt_family="(CA)n"

```

repeatL_region	12437. .12842	/rpt_family="L1"
repeatL_region	12888. .13001	/rpt_family="L1"
repeatL_region	13003. .13358	/rpt_family="MER2"
repeatL_region	13365. .14214	/rpt_family="L1"
repeatL_region	14264. .14417	/rpt_family="L1"
repeatL_region	14415. .14729	/rpt_family="L1"
repeatL_region	14816. .15026	/rpt_family="MER1_type"
repeatL_region	15648. .15729	/rpt_family="CATa)n"
repeatL_region	16041. .16362	/rpt_family="L1"
repeatL_region	16679. .16732	/rpt_family="L1"
repeatL_region	17025. .17315	/rpt_family="Alu"
repeatL_region	17465. .17685	/rpt_family="L1"
repeatL_region	18044. .18633	/rpt_family="L1"
repeatL_region	18630. .18750	/rpt_family="L1"
repeatL_region	18760. .19497	/rpt_family="L1"
repeatL_region	19308. .19692	/rpt_family="L1"
repeatL_region	19702. .19814	/rpt_family="L1"
repeatL_region	21491. .21563	/rpt_family="L2"
repeatL_region	21987. .24424	/rpt_family="L1"
repeatL_region	24425. .24720	/rpt_family="Alu"
repeatL_region	24722. .24821	/rpt_family="L1"
repeatL_region	25383. .25745	/rpt_family="MER1_type"
repeatL_region	29202. .29272	/rpt_family="MER2_type"
repeatL_region	30338. .30386	/rpt_family="CA)n"
repeatL_region	30984. .31045	/rpt_family="MER2_type"
repeatL_region	31041. .31398	/rpt_family="MER2_type"
repeatL_region	32178. .32319	/rpt_family="MER2_type"
repeatL_region	32475. .32515	/rpt_family="L1"
repeatL_region	32589. .32646	/rpt_family="L2"
repeatL_region	32940. .33224	/rpt_family="Alu"
repeatL_region	33843. .33929	/rpt_family="MER1_type"
repeatL_region	35891. .36381	/rpt_family="L2"
repeatL_region	37910. .38123	/rpt_family="MER1_type"
repeatL_region	38302. .38768	/rpt_family="L1"
repeatL_region	42711. .43026	/rpt_family="L1"
repeatL_region	43023. .43418	

```
repeat_region /rpt_family="L1"
43552..43886
/rpt_family="L1"
repeat_region /rpt_family="L1"
43847..44539
/rpt_family="L1"
repeat_region /rpt_family="L1"
44534..45153
/rpt_family="L1"
repeat_region /rpt_family="L1"
45132..45245
/rpt_family="L1"
repeat_region /rpt_family="L1"
45246..45667
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
45674..45869
/rpt_family="L1"
repeat_region /rpt_family="L1"
45871..46089
/rpt_family="Alu"
repeat_region /rpt_family="Alu"
46091..47111
/rpt_family="L1"
repeat_region /rpt_family="L1"
47186..47354
/rpt_family="MER1_type"
repeat_region /rpt_family="MER1_type"
47529..47577
/rpt_family="MER1_type"
repeat_region /rpt_family="MER1_type"
47673..48037
/rpt_family="L2"
repeat_region /rpt_family="L2"
48067..48494
/rpt_family="L2"
repeat_region /rpt_family="L2"
51739..51788
/rpt_family="(CA)n"
repeat_region /rpt_family="(CA)n"
52362..52665
/rpt_family="L1"
```

Query Match	38.8%;	Score 470.2;	DB 9;	Length 97979;
Best Local Similarity	76.8%;	Pred. No. 6e-95;		
Matches 655;	Conservative 0;	Mismatches 158;	Indels 40;	Gaps 5;

OY	329	TCGACATATGAAGGCTGTGGCAAGCCCTTTCATCAGAGACTACACTCTGACGCCACACTT	388
Db	79437	TCGTGACTATATAAAGGGGTGTGTCTAAACCCCTTCATCAGAGACTACCTTCTCAATTCGCCACATC	79378
OY	389	CTGACTCAGCACAGAGAAAAAGCCGTTTGTGTGTGTGAGCAGCTAGGGGTGTATCAAA-----	443
Db	79377	CTGATTCACATTGGAGAAAAAGCCCTTTATTATTGTGTGAGCTAGAGGCTGTGTATAAAAAAA	79318
OY	444	-----AATTCAACACAAATTCAAACTTGAAGAAACATTTTGAAAGCAACATGAAAAATC	497
Db	79317	AAAAAAGATCCACAGAAAAATTCACACTGGAGAAACATTTTTCATGCAAAACATGAAAAATC	79258
OY	498	AACAAAAACAATATATATGCACTTTTGAAGACTGTAAAGAGCTTTTAAAGAACATCAGC	557
Db	79257	TGCAAAAGCAATATGTATGCAATTTTGAAGATGTAAAGAGCTTTTAAAGAGCTTTTCAAGC	79198
OY	558	AGCTGAAAAATCATGAGTCCAGCATACCAATGAACTCTATTCAAAGTACCATCCAGAG	617
Db	79197	AGCTGAAAAATCATGAGTCCAGCATGAAATTAACCCCTATTCAAGTTTAACCTTGAAG	79138
OY	618	GATGTGGGAAACACTTGGCATCACCAGCAGCAAGCTGAAGAGCATGCCAAGGCCACGAG	677
Db	79137	GATGTGGGAAACACTTGGTTTTCACCTCCAGCGTGGAGAGACAGATGCAGGTATTCAGA	79078
OY	678	GCTATGTATGTCAAAAAGGATGTTCTCTTTGTGTGGCAAAACATGACGCACTTCTGAAC	737
Db	79077	AAATATATATGTCAAAAGGATGTTCTTTTGTGTGGCAAAACATGAGC-----	79032
OY	738	ATGTGAGAGAAACCATTAAGAGGAAATAACTGTGAAGTATGCCGGAACCAATTAAAC	797
Db	79031	---AAGAGAAATTCATGAAAAGGAAATA--ACATGATGTATGCCAGAAACATTTCAAT	78977
OY	798	GCAAGATTACCTTAAGCAACATGAAACATCATGCCCAAGAAAGGATGTATGCGT	857
Db	78976	GTGAAGATTATCTCAGATTAACCTGTGAAAACATCATGCCCTCAGAAGAGATGTATGTCAAC	78917
OY	858	GTCCAAGAGAGGCTGTGGAGAACCTATACTACTGTGTTAATCTCCAAACCATATCC	917
Db	78916	ATCCCAAGAGAAACATGCGCAACA-----ACGTATTTTAATCTCCAGAGCTTCTTC	78866

TITLE Direct Submission
JOURNAL Submitted (30-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 111722)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 111722)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jan 19, 2002 this sequence version replaced gl:17976493.

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplons@wustl.wustl.edu

Summary Statistics
Center project name: H_NH0515K14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTD-2105K18, 2000 bp overlap; the clone sequenced to the right is RP11-533K11, 2000 bp overlap. Actual start of this clone is at base position 141753 of CTD-2105K18; actual end is at base position 16829 of RP11-533K11. Polymorphisms have been identified between AC016444 and AC074390. Data from AC016444 was used to finish this clone, AC074390.
Location/Qualifiers
1. 111722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-515K14"

repeat_region /clone_11b="RPCI-11"
1. 249
/rpt_family="MALR"
repeat_region 250. 640
/rpt_family="MALR"
repeat_region 1650. 2014
/rpt_family="ERV1"
repeat_region 2015. 3233
/rpt_family="L1"
repeat_region 3234. 3578
/rpt_family="ERV1"
repeat_region 3586. 3835
/rpt_family="ERV1"
repeat_region 4560. 4873
/rpt_family="ERV1"
repeat_region 4880. 4975
/rpt_family="GA-rich"
repeat_region 4991. 5133
/rpt_family="ERV1"
repeat_region 5577. 5801
/rpt_family="L1"
repeat_region 6687. 6962
/rpt_family="Alu"
repeat_region 7392. 7449
/rpt_family="AT-rich"
repeat_region 7495. 7564
/rpt_family="AT-rich"
repeat_region 7559. 8728
/rpt_family="L1"
repeat_region 8746. 9038
/rpt_family="Alu"
repeat_region 9017. 9053
/rpt_family="(TAA)n"
repeat_region 9061. 9362
/rpt_family="Alu"
repeat_region 9372. 10143
/rpt_family="L1"
repeat_region 9934. 9963
/rpt_family="(TG)n"
repeat_region 10144. 10454
/rpt_family="Alu"
repeat_region 10427. 10454
/rpt_family="(A)n"
repeat_region 10455. 11140
/rpt_family="L1"
repeat_region 11487. 11960
/rpt_family="L2"
repeat_region 12112. 12305
/rpt_family="L2"
repeat_region 12306. 12615
/rpt_family="Alu"
repeat_region 12591. 12615
/rpt_family="(A)n"
repeat_region 12616. 12766
/rpt_family="L2"
repeat_region 13024. 13088
/rpt_family="(TA)n"
repeat_region 13613. 13641
/rpt_family="AT-rich"
repeat_region 14281. 14328
/rpt_family="(TA)n"
repeat_region 14329. 15028
/rpt_family="MER2-type"
repeat_region 15050. 15494
/rpt_family="MER2-type"
repeat_region 15445. 15464
/rpt_family="(T)n"
repeat_region 15495. 16696
/rpt_family="L1"
repeat_region 17544. 17927
/rpt_family="L1"
repeat_region 19584. 19618
/rpt_family="MALR"

```

repeat_region      21646..21769
                    /rpt_family="MIR"
repeat_region      21860..22043
                    /rpt_family="MERL_type"
repeat_region      22044..22427
                    /rpt_family="Mair"
repeat_region      22428..22467
                    /rpt_family="MERL_type"
repeat_region      23300..23664
                    /rpt_family="Mair"
repeat_region      23870..23910
                    /rpt_family="(7AAA)n"
repeat_region      24405..24635
                    /rpt_family="L1"
repeat_region      24636..24901
                    /rpt_family="Mair"
repeat_region      25481..25504
                    /rpt_family="AT_rich"
repeat_region      26543..26672
                    /rpt_family="MERL_type"
repeat_region      26769..26911
                    /rpt_family="CRI"
repeat_region      26928..27295
                    /rpt_family="Mair"
repeat_region      27653..27678
                    /rpt_family="AT_rich"
repeat_region      27934..28093
                    /rpt_family="MERL_type"
repeat_region      29167..29194
                    /rpt_family="AT_rich"
repeat_region      29191..29309
                    /rpt_family="Alu"
repeat_region      29348..29644

```

Query Match 37.38; Score 452.6; DB 9; Length 111722;

Best Local Similarity 80.38; Pired. No. 5.6e-91;

Matches 685; Conservative 0; Mismatches 139; Indels 29; Gaps 12;

```

OY 209 CCGGCGCCCGGCTTCCAGAGGTCATGCTCTCCCTGCTGCGCCCAATTAC 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87601 CCCCTTACACACTTCTTGAAGGGTCTGCTACTTCCC-CAGTGGGACACCTTAC 87543

OY 269 AGCAAGCCTGGAAGCTTGAAGCGCACCCTGTGAAGCAGACAGGGGAGAGACATTGCTT 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87542 AAAAAGCGCTGGAAGCTTAATGTGCACCTGT- AATCAGATGGGGAAGAAACATTGTT 87485

OY 329 TGTACTATGAAGGGTGTGCAAGCCTTCTATC-AGGACTACATCTGAGCCGCACAT 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87484 TATGACTGTGAAGGGTGTGCGCAAGACCCACCTCAAGAGACTAACATCTGAGTTGTCAATGT 87425

OY 388 TCTGACTCACAGAGAGAAAGCGGTTTGTGTGCAAGCCACTGCTGTGATC-AAAAAT 446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87424 CCTGA-TCATACCTGACAGCAAGCCATATGTTGTGCAAGCTAGTACTGTCTTAATAAT 87366

OY 447 TCAACACAAATCAAACTTGAAGAAACATTTTGAAGCAACATGAAATACACAAAAA 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87365 TCAACACAAATCAAACTTGAAGAGACATTTTACAGGCAACATGAAACACAGAAAAA 87306

OY 507 A-----ATATATGCAAGTTTGAAGACTGTAAAGACCTTAAAGAAACATGACACCT 561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87305 AAAAATATGATTAATGCAATTTTGAAGGTGGAAGAAAGCTTTTACG-AACATCAGCAGCT 87247

OY 562 GAAATCATGAGTGCAGCATACCAATGAACCTCTATTAAGTGTACCCAGGAAGATG 621
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87246 GAAAAACCATCTAGTCCAGCATACCAATGAACCATATTCAACTGTACCCAGGAAGATG 87187

OY 622 TGGGAACACTTGTGCATCCAGCAAGCTGAAACAGCATGCCAAGGCCAAGAGGCTA 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87186 TGGGAAGCATTTTCCCATCATCCAGC-----AAGCGCAAGGCCCATGCGGGCTC 87134

OY 682 TGTATGTCAAAAAAGATGTCTCTTTGTGGCAAAAAACATGACGAACTTGTGAACATGT 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87133 TATATGTCAAAAAAGCGTGTCTTTTGTGTGAAGCATGAAGAAAGTTTGAACATGT 87074

```

```

OY 742 GAGAGAAACCCATAAAGAGAAATACTATGTGAAGTATGCCGGAACATTTAAGCAA 801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87073 GAGAGAAACCTTATTAAGAAAGCAACATATGAAGCGGCGAGAAACATTTTAAC---- 87018

OY 802 AGATTACCTTAAGCAGACATGAAACTCATGCCCGAAGAGGATGTATGCTGTCC 861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87017 -ATTATCTTAAGCAACATATGAACACTCATGCCCGAAGAGGATGT-TTTTGACATAC 86961

OY 862 AAGGAGGCTGTGGAAGAACCTACTACTGTCTTAATCTCCAAAGCATATCCCTGC 921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86960 AAGGAGGCTCTGGAAGAACCTACACACTGTCTTACTCTCAAAATCATATCTTTC 86901

OY 922 CTTTCATGAGGAAGCCGCCCTTTTGTGTGAACATGTGCTGGCAAAACATTTGC 981
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86900 CAT--GAGAAAAAAGGTGCCATTTTGCATGTGAACATGTGGCTGTGGCAAAACATTTGC 86843

OY 982 AATGAACAAAGCTCTCAAGGCAAGCTGTGTATCATGATCTGACAAAGAAATGAA 1041
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86842 AATGAACAAAGCTCTTCTAGGCAATGCTGTGTACAT-ATTCTGACAAAGAAATGAA 86784

OY 1042 GCTCAAGTCAAA 1054
    ||||| ||||| |||||
Db 86783 GCTCAGAGTATA 86771

```

```

RESULT 12
AC016444      206622 bp      DNA      linear      HTG 04-MAY-2001
LOCUS
AC016444
DEFINITION
Homo sapiens chromosome 11 clone RP11-384114 map 11, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
AC016444
AC016444.3  GI:10045722
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 206622)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-384114
Unpublished
2 (bases 1 to 206622)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deatellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hages,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwen,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,M.J., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
COMMENT
Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6778507.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTB
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L5006
Center clone name: 384_1_14
-----
Summary Statistics

```



```

||||| 11 ||||||| 11 ||||||| 11 ||| 11 ||||||| 11
Db 36949 AAGAGAGTCTGTGGAGAACCTACACAGTGTACTTACTTACAAATCATATCTTTC 37008
QY 922 CTTCCATGAGGAAGACCGCCCTTTTGTGTGAAACATGCTGCGTGGCAAAATTTGC 981
Db 37009 CAT--GAGAAAAAGGTGCTATTTTGCATGTGGTGCATGTCGCAAAACATTTGC 37066
QY 982 AATGAACAAGTCTCATGAGCATGCTGTGTACATGATCCGTGACAAAGAAATGAA 1041
Db 37067 AATGAACAAGTCTTCTAGGATGCTGTGTATCAT--ATTCTGACAAAGAAATGAA 37125
QY 1042 GCTCAAGTCAAA 1054
Db 37126 GCTCAGAGTAATA 37138

RESULT 13
AL353741 199517 bp DNA linear PRI 28-JAN-2001
LOCUS Human DNA sequence from clone RP11-575C20 on chromosome 9, complete
DEFINITION sequence.
ACCESSION AL353741
VERSION AL353741.16 GI:12584694
KEYWORDS RTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 199517)
AUTHORS Laird,G.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
on Jan 28, 2001 this sequence version replaced gi:12580980.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-575C20 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-575C20 The true
left end of clone RP11-279E1 is at 116738 in this sequence. The
true right end of clone RP11-21817 is at 84470 in this sequence.
FEATURES
source
location/Qualifiers
1..199517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-575C20"
/clone_1lb="RPCI-11.2"
159090..159136
/misc_feature
/note="Single clone region. Assembly confirmed by
restriction digest data."
159357..159427
misc_feature

```

```

/note="Single clone region. Assembly confirmed by
restriction digest data."
BASE COUNT 61043 a 40153 c 38840 g 59481 t
ORIGIN
Query Match 36.9%; Score 447; DB 9; Length 199517;
Best Local Similarity 75.1%; Pred. No. 1.1e-89;
Matches 725; Conservative 0; Mismatches 180; Indels 60; Gaps 11;
QY 131 GTGTCCCGAGTGTGTGCTCTTGTACACATGCCGACGCTTCAATGACCGCGAG 190
Db 123892 GTGGGTCCCGAGGCAATGTCACCCCTGAGCATTCCTAAGCATTTTCGTGCGCCGTAG 123951
QY 191 AGCTCAGCTCC-----GACCCCGCGCGCGCGCGCTCCAGAGGATTCATGCT 242
Db 123952 AGACCACTCCACCTCCCTTGGTCCCTCTTCCCTGTGACCTCTTAAGATTTCACTGTCT 124011
QY 243 CCTTCCTGACTGC-----AGCGCAATTACAGCAAAAGCTTGAGAGTTCAGCGCACCTG 298
Db 124012 CCTTCAGAGACTGTATGTAGACACCAATTACATTAAGTCCAGATGCTTGAAGAGCACCTA 124071
QY 299 TGCAGACACACGGGGAGAGACCATTTGTTGTGACTATGCAAGGCTGTGCCAAGCCTTC 358
Db 124072 TGCAGACACACGAGGAAGAGACCATTTGTTGTGACTATGCAAGGCTGTGCCAAGATTTTC 124131
QY 359 ATCAGGACACTACATCTGAGCGCGCACATCTGTGACACAGGAGAAAGCCTTTGTT 418
Db 124132 CTCAGGGTCTACACCTGAGACACACATCTGTGATATATGAGAAAGCTTTTAT- 124190
QY 419 TGTGACACCACTGCTGTGATCAAAAATTCAACCAAAATCAACTGAAAGAAACATTTT 478
Db 124191 -----GCACTAGTGGCTGTGACAGAAATTTTCAACCAAAATCAACTGAAAGAAATTTT 124246
QY 479 GAACGCAACATGAAATCAACAAACAAATATATATGACATTTTGAAGACTGTAAGAG 538
Db 124247 GAACGCAACATGAAATGATTCAGC--AAACACATATATGCAATTTTGAAGGCTGTAAGAAAG 124304
QY 539 ACCTTTAAAGAACTTCAGACCTGAAATTCATAGCGCCAGCATACCAATGAACCTCTA 598
Db 124305 ATCTTTAAG--GATCTAGTACTGAAATTCATAGTCCAGCTACCAATGA----- 124354
QY 599 TTCAAGTGTACCCAGGAAGATGTGGGAAACACTTTCATCCACCAGCAAGCTGAAAGCA 658
Db 124355 -----ACCCAGAAAGAGTGTGGGAATAC--TGGCTGATCCAGAGGCTGAAAGCA 124403
QY 659 CATGCCAAGGCCCGACGAGGCTATGTATGTCAAAAAGATTTCTTTGGCAAAAACA 718
Db 124404 CACACGAGAGGTTCAAGAGGCGCTA--GATGTCAAAAAGATGTTCTTTGTGGCAAAAACA 124461
QY 719 TGGACGGAACCTTGGAAACATGTGAG-----AGAAACCATTAAGAGAAATATATGTA 774
Db 124462 TGGGTTGAATCTTGAAATATGTATGTAGAAABAAACCATATATAGAGAAATATACATGTGA 124521
QY 775 AGTATGCGGGAATCATTTAAACGCAAAAGATTACCTTAAGCAACATGAAACATCATGC 834
Db 124522 A-TATGGAGAAATTTTAAACCAAAAGATTATCTTAAGCAATATATGAAACTCTTCC 124580
QY 835 CCCGAAAGGATATATGTGCTGTCCAGAGAAAGCTGTGGAAGAACTATACTGT 894
Db 124581 TCCAGAGGAATATTTGCCAATGTTCAGAGAGAGGCTGTGTGAACCCACATGACTGT 124640
QY 895 GTTTAATCTCCAAAGCATATATCTCTCTCTCATGAGAGAAAGCGCCCTTTGTGTGTA 954
Db 124641 ATTTAATCTACAGAGCTGTATCTCTCTTTCAATGAGAGAAAGGACCATATTTGTGTGTA 124700
QY 955 ACATGCTGCTGTGCAAAACATTTGCAATGAAACAAAGTCTCATAGGCAATGCTTTGT 1014
Db 124701 CCATCTGCGCTGTGCAAAAGAGTGTGCAATGAA-----AGTCAATATGAC 124747
QY 1015 ACATGATCTCTGCAAGAAAGAAATGAAGCTCAAAAGTCAAAATATCTGCTGTAAGAGGAG 1074
Db 124748 ATACATTTGCAACAAAGAAATTAATTAATCAGAGTAAACCAATCATGAAAAACGAG 124807

```

QY 1075 TTTCG 1079
1111
Db 124808 TTCCG 124812

RESULT 14
AX396111/c AX396111 439 bp DNA linear PAT 18-MAY-2002
LOCUS Sequence 326 from Patent WO0212328.
DEFINITION AX396111
ACCESSION AX396111
VERSION AX396111.1 GI:21066858
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS King, G.E., Meagher, M.J., Xu, J., and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A 326 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
Source Location/Qualifiers
1. .439
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 106 a 94 c 99 g 140 t
ORIGIN

Query Match 34.5%: Score 418.4: DB 6: Length 439:
Best Local Similarity 99.8%: Pred. No. 1.1e-83:
Matches 419: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 794 AAAGCGAAGATTACCTTAAGCAACATGAAACATGATGCCCGAAGAGATGTATGT 853
|||||
Db 439 AAAGCGAAGATTACCTTAAGCAACATGAAACATGATGCCCGAAGAGATGTATGT 380
QY 854 CGCTGTCCAAAGAGAGCGCTGTGGAAGACCTATACCTGTTTAAATCTCCAAAGCCAT 913
|||||
Db 379 CGCTGTCCAAAGAGAGCGCTGTGGAAGACCTATACCTGTTTAAATCTCCAAAGCCAT 320
QY 914 ATCTCTCTCCATGAGGAAGCGCCCTTTGTGTGTAACATGCTGGCTGTGGCAA 973
|||||
Db 319 ATCTCTCTCCATGAGGAAGCGCCCTTTGTGTGTAACATGCTGGCTGTGGCAA 260
QY 974 ACATTTGCATGAAACAAATCTCTACCTAGCGATGCTTTGATGATCTCTGACAAAG 1033
|||||
Db 259 ACATTTGCATGAAACAAATCTCTACCTAGCGATGCTTTGATGATCTCTGACAAAG 200
QY 1034 AAATGAAAGCTCAAAAGTCAAAAATCTCGTGAAGAAAGGAGTTGGCTCTCATCTCAGT 1093
|||||
Db 199 AAATGAAAGCTCAAAAGTCAAAAATCTCGTGAAGAAAGGAGTTGGCTCTCATCTCAGT 140
QY 1094 GGAATATATCCCTCCCAAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAAGGAGAG 1153
|||||
Db 139 GGAATATATCCCTCCCAAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAAGGAGAG 80
QY 1154 TCACCCCAACTGTGTGGAAGACAGATGCTCTCGACACTTGCAGTACTTACCTTTGGCTAA 1213
|||||
Db 79 TCACCCCAACTGTGTGGAAGACAGATGCTCTCGACACTTGCAGTACTTACCTTTGGCTAA 20

RESULT 15
AX396474 439 bp DNA linear PAT 18-MAY-2002
LOCUS Sequence 689 from Patent WO0212328.
DEFINITION AX396474
ACCESSION AX396474
VERSION AX396474.1 GI:21067221
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1
AUTHORS King, G.E., Meagher, M.J., Xu, J., and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A 689 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
Source Location/Qualifiers
1. .439
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 141 a 99 c 93 g 106 t
ORIGIN

Query Match 34.4%: Score 416.8: DB 6: Length 439:
Best Local Similarity 99.5%: Pred. No. 2.6e-83:
Matches 418: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

QY 794 AAAGCGAAGATTACCTTAAGCAACATGAAACATGATGCCCGAAGAGATGTATGT 853
|||||
Db 1 AAAGCGAAGATTACCTTAAGCAACATGAAACATGATGCCCGAAGAGATGTATGT 60
QY 854 CGCTGTCCAAAGAGAGCGCTGTGGAAGACCTATACCTGTTTAAATCTCCAAAGCCAT 913
|||||
Db 61 CGCTGTCCAAAGAGAGCGCTGTGGAAGACCTATACCTGTTTAAATCTCCAAAGCCAT 120
QY 914 ATCTCTCTCCATGAGGAAGCGCCCTTTGTGTGTAACATGCTGGCTGTGGCAA 973
|||||
Db 121 ATCTCTCTCCATGAGGAAGCGCCCTTTGTGTGTAACATGCTGGCTGTGGCAA 180
QY 974 ACATTTGCATGAAACAAATCTCTACCTAGCGATGCTTTGATGATCTCTGACAAAG 1033
|||||
Db 181 ACATTTGCATGAAACAAATCTCTACCTAGCGATGCTTTGATGATCTCTGACAAAG 240
QY 1034 AAATGAAAGCTCAAAAGTCAAAAATCTCGTGAAGAAAGGAGTTGGCTCTCATCTCAGT 1093
|||||
Db 241 AAATGAAAGCTCAAAAGTCAAAAATCTCGTGAAGAAAGGAGTTGGCTCTCATCTCAGT 300
QY 1094 GGAATATATCCCTCCCAAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAAGGAGAG 1153
|||||
Db 301 GGAATATATCCCTCCCAAAAGCAAGGCAAGGCTTATCTTTGTGTCAAAAGGAGAG 360
QY 1154 TCACCCCAACTGTGTGGAAGACAGATGCTCTCGACACTTGCAGTACTTACCTTTGGCTAA 1213
|||||
Db 361 TCACCCCAACTGTGTGGAAGACAGATGCTCTCGACACTTGCAGTACTTACCTTTGGCTAA 420

Search completed: February 10, 2003, 17:02:32
Job time: 3923.82 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:56:36 ; Search time 1896.59 Seconds
(Without alignments)
10356.112 Million cell updates/sec

Title: US-09-831-426C-4
Perfect score: 1213
Sequence: 1 gtccgcgcgcgcgcgaag.....cagtaactacctgtgctaa 1213

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estum:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrl:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	921.4	76.0	1078	13	BM553401
2	919	75.8	958	9	AL518337
3	904	74.5	916	9	AL550467
4	877.8	72.4	881	9	AL555777
5	870.6	71.8	1059	13	BM562682
6	862.8	71.1	937	9	AL574318

Result No.	Score	Query Match	Length	DB ID	Description
7	860.4	70.9	930	14	BO650498
8	849.4	70.0	957	9	AL527493
9	844.6	69.6	906	9	AL574698
10	838.8	69.2	893	9	AU117724
11	836	68.9	902	14	BO644275
12	830.2	68.4	933	9	AL578193
13	829	68.3	1089	14	BO278693
14	827.4	68.2	927	12	BO330080
15	827.2	68.2	908	14	BO648620
16	824	67.9	868	12	BO683119
17	822.6	67.8	1110	14	BM809285
18	818.6	67.5	1111	13	BM541851
19	817.4	67.4	892	9	AL518336
20	812	66.9	871	12	BO750799
21	810.4	66.8	900	12	BO697646
22	806.8	66.5	899	14	BO332064
23	801.4	66.1	971	14	BO279125
24	800.4	66.0	842	14	BO430441
25	799	65.9	887	14	BO225045
26	797.2	65.7	880	13	BO47210
27	796.2	65.6	912	12	BE797173
28	796	65.5	910	12	BO281954
29	792.6	65.3	974	13	BO86290
30	791.2	65.2	1027	10	BE562240
31	787.6	64.9	932	9	AL527492
32	784	64.6	883	14	BO436552
33	783.8	64.6	923	13	BM453967
34	781.6	64.4	939	14	BO293358
35	778.8	64.2	879	13	BM041009
36	778	64.1	953	13	BM041678
37	774.8	63.9	801	12	BM55545
38	768.4	63.3	1047	13	BM55545
39	763	62.9	1043	12	BO389788
40	750.4	61.9	874	10	BE410109
41	749.8	61.8	755	13	BE410109
42	738.6	60.9	878	12	BO822097
43	737.8	60.8	753	13	BI759103
44	736.2	60.7	839	13	BI598078
45	732.8	60.4	736	13	BI764623

ALIGNMENTS

RESULT 1
LOCUS BM553401 1078 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6572564 NIH_MGC_41 Homo sapiens CDNA clone IMAGE:5467230
ACCESSION BM553401
VERSION BM553401.1 GI:18792097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1078)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-f@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
plate: LCM1968 row: 0 column: 07
High quality sequence stop: 673.
Location/Qualifiers 1..1078

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5467230"
/clone_1lb="NH_MGC_41"
/lssue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pOTB1; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
SuperScript II RT (Life technologies). Note: this is a
NH_MGC Library."
BASE COUNT      301 a      276 c      275 g      223 t      3 others
ORIGIN
Query Match      76.0% Score 921.4; DB 13; Length 1078;
Best Local Similarity 95.1%; Pred. No. 5.9e-194;
Matches 1025; Conservative 0; Mismatches 44; Indels 9; Gaps 7;

QY      6  GCGCCGCGCGAAGCTTCAGAGGAGCCGTGGCCGCGCGCGGTTCCCGGACGTG 65
      1  GCGCTCGCGGAAGCTTCAGAGGAGCCGTGGCCGCGCGCGGTTCCCGGACGTG 60
QY      66  TCTCGGACGTGGCAGCGCCCTGGCCCTGGCTTGGAGCGCGCGCGCTGGATCCG 125
      61  TCTCGGACGTGGCAGCGCGCCCTGGCCCTGGCTTGGAGCGCGCGCGCTGGATCCG 120
QY      126  GCGCCGTGGTCCGAGTCCGTGCTCTTGAACCATCCCGACGCTTCATTCAGCCG 185
      121  GCGCCGTGGTCCGAGTCCGTGCTCTTGAACCATCCCGACGCTTCATTCAGCCG 180
QY      186  GCGAGACTAGCTCGAGCCCGCGCGCGCTTCCAGAGGTATATGCTGCT 245
      181  GCGAGACTAGCTCGAGCCCGCGCGCGCTTCCAGAGGTATATGCTGCT 240
QY      246  TCCCTGACTGACGCCAATTACAGCAAGCTTGAAGCTTGAAGCTTGAAGC 305
      241  TCCCTGACTGACGCCAATTACAGCAAGCTTGAAGCTTGAAGCTTGAAGC 300
QY      306  ACAGCGGGGAGAGACATTTGTTGTGACTATGAAGGTGGCAGAGCCCTCATCAGG 365
      301  ACAGCGGGGAGAGACATTTGTTGTGACTATGAAGGTGGCAGAGCCCTCATCAGG 360
QY      366  ACTACATCTGAGCGGCACATTTCTGACTACAGAGGAGAAAGCCGTTGTTGTGAG 425
      361  ACTACATCTGAGCGGCACATTTCTGACTACAGAGGAGAAAGCCGTTGTTGTGAG 420
QY      426  CCAGTGGCTGTATCAAAATTTCAACACAAATCAAACTTGAAGAACGCA 485
      421  CCAGTGGCTGTATCAAAATTTCAACACAAATCAAACTTGAAGAACGCA 480
QY      486  AACATGAAATCAACAAACATATATGACAGTTTGAAGCTTGAAGACCTTTA 545
      481  AACATGAAATCAACAAACATATATGACAGTTTGAAGCTTGAAGACCTTTA 540
QY      546  AGAATCATGACGCTGAAATCCATGAGTGGCAGATACCAATGAACCTTATCAAGT 605
      541  AGAATCATGACGCTGAAATCCATGAGTGGCAGATACCAATGAACCTTATCAAGT 600
QY      606  GTACCCAGAGAGATGTGGGAAACATTTGCATCACCAGCAAGCTGAAGACATCCCA 665
      601  GTACCCAGAGAGATGTGGGAAACATTTGCATCACCAGCAAGCTGAAGACATCCCA 660
QY      666  AGGCCACAGAGAGGCTATGTATGTCAAAAGAGATGTTCTTTGGGCAAAAACATGACG 725
      661  AGGCCACAGAGAGGCTATGTATGTCAAAAGAGATGTTCTTTGGGCAAAAACATGACG 720
QY      726  AACTTGAAGACATGTAGAGAAACCATTAAGAGAAATATCTATGTGAAGTATGCCGGA 785
      721  AACTTGTG-AACATGTAGAGAAACCATTAAGAGAAATATCTATGTGAAGTATGCCGGA 779

```

```

QY      786  AACATTTAAACGCAAGATTACCTTAAGCAACATGAAACATCATGCCCAAGAAAGG 845
      780  NAACATTTAAACGCAAGATTACCTTAAGCAACATGAAACATCATGCCCAAGAAAGG 838
QY      846  ATGTATGTGCTGTCCAGAGAGAGCTGTGAGAGACCTTACTACTGTGTTTATCTCC 905
      839  ATGTATGTGCTGTCCAGAGAGAGCTGTGAGAGACCTTACTACTGTGTTTATCTCC 898
QY      906  AAACCATATCCCTCCCTTCATGAGAGAA-CGCCGCTTTTGTGTGAACATGCT--G 962
      899  -AAACCATATCCCTCCCTTCATGAGAGAAAGCGGCTTTGTGTGAACATGCTTTGG 957
QY      963  GCTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGGATGCTGTTGTA--CATGA 1020
      958  CTGTGGGCAAAACATTTGCAATGAACAAAGTCTCCCTAGGATGCTGTTGTAACATGAT 1017
QY      1021  TCTGTACAGAGAAAT--GAAGCTCAAAAGTCAAAAATCTGTGAAAACGAGTTT 1077
      1018  CCCTGACAGAGAAATGGAACCTTCAAAATTTCCGAAATTTCCGAAAAACGAGATT 1075

RESULT 2
AL518337
LOCUS      AL518337 LTI_NFLO11_NBC1 Homo sapiens cDNA clone CS0DA009Y018 5
DEFINITION prime, mRNA sequence.
ACCESSION AL518337
VERSION    AL518337.1 GI:12781830
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 958)
            J.L.W.B., Gruber,C., Jesse,J. and Polayes,D.
            Full-length cDNA libraries and normalization
            JOURNAL
            Unpublished (2001)
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES
            source
            1..958
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DA009Y018"
            /clone_1lb="LTI_NFLO11_NBC1"
            /sex="male"
            /lssue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            Eco RV sites of the pCMVSPORT 6
            vector. Library is not normalized, but is the control for
            the normalized libraries. Library was constructed by Life
            Technologies. Contact : Feng Liang Life Technologies, a
            division of Invitrogen 9800 Medical Center Drive Rockville
            , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
            fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT      270 a      244 c      232 g      195 t      17 others
ORIGIN
Query Match      75.8% Score 919; DB 9; Length 958;
Best Local Similarity 97.5%; Pred. No. 2e-193;
Matches 935; Conservative 16; Mismatches 6; Indels 2; Gaps 2;

QY      16  GAAGCTCAGCAGGAGCCGCGCGCGCGGTTCCCGGACGTGCTCGGACG 75
      1  GAAGCTTACCAGGAGCCGCGCGCGCGGTTCCCGGACGTGCTCGGACG 60
QY      76  TCGCAGCGCGCTGCGCTTGGAGCGCGCGCGCTGTGATCCGCGCGCTGCT 135

```



```

|||||
Db 61 TGGCAGCGCCGCTGGCCCTGGGCTTGGAGGCGCGGCCCTTGATCCCGCGCGTGGT 120
136 CGCGGAGTGGGTGTCGCTTGGACATGGCGAGCGGTCAATGACGCGCGAGAGCTC 195
121 CGCGGAGTGGGTGTCGCTTGGACATGGCGAGCGGTCAATGAGCGCGAGAGCTC 180
196 AGCTCCGACCGCGCGCGCGCGCTTCCAGAGAGTTCATGCTGCTTCCCTGACTG 255
181 AGCTCCGACCGCGCGCGCGCGCTTCCAGAGAGTTCATGCTGCTTCCCTGACTG 240
256 CAGCGCCCAATTACAGCAAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 315
241 CAGCGCCCAATTACAGCAAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 300
316 GAGACCATTTGTTGTCATGTAGAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAG 375
301 RARACCATTTGTTGTCATGTAGAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAG 360
376 GAGCGCGCACATTCATGTCACAGAGAGAGCGCGCTTGTGTGTGCGAGCGCTGCTG 435
361 RAGCGCGCACATTCATGTCACAGAGAGAGCGCGCTTGTGTGTGCGAGCGCTGCTG 420
436 TGATCAAAAATTCAACACAAAATCAAACTGAAGAACTTTTGAACGCAACATGAAA 495
421 TGATCAAAAATTCAACACAAAATCAAACTGAAGAACTTTTGAACGCAACATGAAA 480
496 TCAACAAAACAAATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 555
481 TCAACAAAACAAATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 540
556 GCACCTGAAAATCCATCAGTCCAGCATACCAATGAACCTTATTCAGTGTACCCAGA 615
541 GCACCTGAAAATCCATCAGTCCAGCATACCAATGAACCTTATTCAGTGTACCCAGA 600
616 AGATGTGGGGAAGACCTTGGATCAACCCAGAGCTGAAGAGATGCAAGCGCCAGCA 675
601 AGATGTGGGGAAGACCTTGGATCAACCCAGAGCTGAAGAGATGCAAGCGCCAGCA 660
676 GGGGTATGTATGCAAAAAGATGTCCTTGTGTGGCAAAAACATGAGAGCACTCTAA 735
661 GGGGTATGTATGCAAAAAGATGTCCTTGTGTGGCAAAAACATGAGAGCACTCTAA 720
736 ACATGTGAGAGAAACCCATAAGAGAAATCTATGTAGATGATGATGATGATGATGAT 795
721 ACATGTGAGAGAAACCCATAAGAGAAATCTATGTAGATGATGATGATGATGATGAT 780
796 AGCAGAGATTACCTTAAAGCAACATGAAAACCTGATGATGATGATGATGATGATG 854
781 AGCAGAGATTACCTTAAAGCAACATGAAAACCTGATGATGATGATGATGATGATG 840
855 GCTGTCCAAAGAGAGGCTGTGAAGAACCTTACTGCTGTTTAAATGCCAAAGCCATA 914
841 GCTGTCCAAAGAGAGGCTGTGAAGAACCTTACTGCTGTTTAAATGCCAAAGCCATA 900
915 TCTCTCTCTTCATGAGAAAGCGCGCTTGTGTGTGCAATGCTGCTGTGGCAAA 973
901 TCTCTCTCTTCATGAGAAAGCGCGCTTGTGTGTGCAATGCTGCTGTGGCAAA 958

```

```

AUTHORS L.I.W.B., Gruber,C., Jesse,J., and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="LFI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; site:1; NotI: 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@life-tech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 261 a 238 c 225 g 190 t 2 others
ORIGIN
Query Match 74.5% Score 904; DB 9; Length 916;
Best Local Similarity 99.6% Pred. No. 4.2e-10;
Matches 904; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 48 GCCGGTTCGCGACAGTGTCTCGGACAGTGGCAGCGCGCTGGGCTTGGAGGCG 107
Db 1 GCCGGTTCGCGACAGTGTCTCGGACAGTGGCAGCGCGCTGGGCTTGGAGGCG 60
QY 108 CCGGCGCCCTGGATCCGCGCGCGCGTGTGCGCGAGTGTGCTTGTGACCATGCGC 167
Db 61 CCGGCGCCCTGGATCCGCGCGCGCGTGTGCGCGAGTGTGCTTGTGACCATGCGC 120
QY 168 ACGCTTATTGACAGCGCGCGAGAGCTCAGTCCGACCCCGCGCGCGCTTCCCA 227
Db 121 ACGCTTATTGACAGCGCGCGAGAGCTCAGTCCGACCCCGCGCGCGCTTCCCA 180
QY 228 GGAGTTCATGTCGCTCTTCCTGCTGCTGACGAGCGCAATTAACAGCAACCTGGAAGCTTG 287
Db 181 GGAGTTCATGTCGCTCTTCCTGCTGCTGACGAGCGCAATTAACAGCAACCTGGAAGCTTG 240
QY 288 ACGCGACCTGTGCAAGCAGACGCGGGGAGAGACATTTGTTGTGATGATGAGGCTGTG 347
Db 241 ACGCGACCTGTGCAAGCAGACGCGGGGAGAGACATTTGTTGTGATGATGAGGCTGTG 300
QY 348 GCAAGCGCTTCATCGAGGAGTACCATCTGAGCGCGCAATTTCTGACTACACAGAGAAA 407
Db 301 GCAAGCGCTTCATCGAGGAGTACCATCTGAGCGCGCAATTTCTGACTACACAGAGAAA 360
QY 408 AGCCGTTTGTGTGTCACAGCACTGCTGTGATCAAAAATTCAACACCAAAATCAAACTTGA 467
Db 361 AGCCGTTTGTGTGTCACAGCACTGCTGTGATCAAAAATTCAACACCAAAATCAAACTTGA 420
QY 468 AGAAACATTTTGAACGCAACATGAAAATCAACAAAACATATATATGACATTTTGAAG 527
Db 421 AGAAACATTTTGAACGCAACATGAAAATCAACAAAACATATATATGACATTTTGAAG 480
QY 528 ACTGTAAAGAACCTTTAAAGAAACATGACAGCTAAATTCATATGAGGCCAGCTATACA 587
Db 481 ACTGTAAAGAACCTTTAAAGAAACATGACAGCTAAATTCATATGAGGCCAGCTATACA 540
QY 588 ATGAACCTCTATTCAGAGTACCCGAGAGATGGGAAACACTTTGCTATCACCCAGCA 647
Db 541 ATGAACCTCTATTCAGAGTACCCGAGAGATGGGAAACACTTTGCTATCACCCAGCA 600
QY 648 AGCTGAAACGACATGCCAAGGCCACGAGGGCTATGATGTCAAAAAGAGTTCCTTTG 707

```

Db 601 AGCTGAACACATGCCAGAGCCGACGAGGGCTATGTATGTCAAAAAAGATGTTCTTTG 660

Qy 708 TGGCAAAAACATGAGCAGCACTTCTGAACATGTGAGAGAAACCATTAAGAGGAATAC 767

Db 661 TGGCAAAAACATGAGCAGCACTTCTGAACATGTGAGAGAAACCATTAAGAGGAATAC 720

Qy 768 TATGTGAATGTGCGGAGAAACATTTAAAGCGAAAGATTACCTTAAGCAACATGAAAA 827

Db 721 TATGTGAATGTGCGGAGAAACATTTAAAGCGAAAGATTACCTTAAGCAACATGAAAA 780

Qy 828 CTATATGCCCCAGAAAGGATGTATGTGCTGTCTCCAGAGAAAGGCTGTGAGAAACCTATA 887

Db 781 CTATATGCCCCAGAAAGGATGTATGTGCTGTCTCCAGAGAAAGGCTGTGAGAAACCTATA 840

Qy 888 CTATGTGTTTAATCTCCAAAGCCATATCTCTCTCCATGAGAGAAAGCGCCCTTTG 947

Db 841 CAAGTGTGTTTAATCTCCAAAGCCATATCTCTCTCCATGAGAGAAAGCGCCCTTTG 900

Qy 948 TGTGTGAA 955

Db 901 TGTGTGAA 908

RESULT 4
AL555777 881 bp mRNA linear EST 16-FEB-2001
LOCUS AL555777 LTI_NFL006.Pl2 Homo sapiens cDNA clone CS0DK002YJ18 5
DEFINITION prime, mRNA sequence.
ACCESSION AL555777
VERSION AL555777.1 GI:12897829
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 881)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL006.Pl2"
/clone_id="CS0DK002YJ18"
/issue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 252 a 229 c 225 g 175 t

ORIGIN

Query Match 72.4%; Score 877.8; DB 9; length 881;
Best local similarity 99.8%; Pred. No. 2.7e-184;
Matches 879; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 22 TCAGCAGGAGCGGTGGCGCGGCGCGGTCGCCGCACTGTCTCGCAGCTGGCAG 81

Db 1 TCAGCAGGAGCGGTGGCGCGGCGCGGTCGCCGCACTGTCTCGCAGCTGGCAG 60

Qy 82 CGGCGCTGGCGCTGGCGCTGGAGGCGCGCGCTGGATCCCGCGCGTGGTCCGCA 141

Db 61 CGCGCTTGCCCTTGCGCTTGAGGCGCGCGGCCCTGGATCCGCGGCGGTGGCGCA 120

Qy 142 GTGCGTGTGCTGCTTGTGACATGCGCCGAGCGGTTATTTGACAGCCGCGAGAGCTACGTC 201

Db 121 GTGCGTGTGCTGCTTGTGACATGCGCCGAGCGGTTATTTGACAGCCGCGAGAGCTACGTC 180

Qy 202 GACCCCGCGCGCGCGCGCTTCCAGAGAGTTGATGCTGCTTCCCTGACTGACGCG 261

Db 181 GACCCCGCGCGCGCGCGCTTCCAGAGAGTTGATGCTGCTTCCCTGACTGACGCG 240

Qy 262 CAATTACAGCAAAAGCTTGAAAGCTTGACGCGACCTTGCAAGCAGCAGGCGGAGAGACC 321

Db 241 CAATTACAGCAAAAGCTTGAAAGCTTGACGCGACCTTGCAAGCAGCAGGCGGAGAGACC 300

Qy 322 ATTTGTTTGACATATGAAAGGTTGGCAAGGCTTATCATGAGGACTATCATCTGACGCG 381

Db 301 ATTTGTTTGACATATGAAAGGTTGGCAAGGCTTATCATGAGGACTATCATCTGACGCG 360

Qy 382 CCACATCTGACATCAGCAGAGCAAAAGCGGTTGTTGTCAGCAGCACTGCTGATCA 441

Db 361 CCACATCTGACATCAGCAGAGCAAAAGCGGTTGTTGTCAGCAGCACTGCTGATCA 420

Qy 442 AAAATTCAGCAAAATCAAACTTGAAGAAACATTTTGAACGCAACATGAAAAATCAACA 501

Db 421 AAAATTCAGCAAAATCAAACTTGAAGAAACATTTTGAACGCAACATGAAAAATCAACA 480

Qy 502 AAAACATATATATGCAAGTTTGAAGACTGTAGAAACCTTTTAAGAAACATCAGCAGCT 561

Db 481 AAAACATATATATGCAAGTTTGAAGACTGTAGAAACCTTTTAAGAAACATCAGCAGCT 540

Qy 562 GAAATCCATCAGTGCAGCAGATACCAATGAACCTCTATCAAGTGTACCCAGAGAGATG 621

Db 541 GAAATCCATCAGTGCAGCAGATACCAATGAACCTCTATCAAGTGTACCCAGAGAGATG 600

Qy 622 TGGCAAAACATTTGCAATCACCAGCAAGCTGAAAGCAGATGCCAGGCCAGAGGCTA 681

Db 601 TGGCAAAACATTTGCAATCACCAGCAAGCTGAAAGCAGATGCCAGGCCAGAGGCTA 660

Qy 682 TGTATGCAAAAAGGATGTTCTTGTGGCAAAAACATGACGGAACCTTGCAAAACATGT 741

Db 661 TGTATGCAAAAAGGATGTTCTTGTGGCAAAAACATGACGGAACCTTGCAAAACATGT 720

Qy 742 GAGCAAAACCCATTAAGAGGAAATACTATGTGAAGTATGCGGAAACATTTAAACGCA 801

Db 721 GAGCAAAACCCATTAAGAGGAAATACTATGTGAAGTATGCGGAAACATTTAAACGCA 780

Qy 802 AGATTACCTTAAGCAGACATGAAGAACTATGCCCGCAAGAGGATGTATGCGTGC 861

Db 781 AGATTACCTTAAGCAGACATGAAGAACTATGCCCGCAAGAGGATGTATGCGTGC 840

Qy 862 AAGGAAGGCTGTGGAAGAACTATACTGTTTAAATC 902

Db 841 AAGGAAGGCTGTGGAAGAACTATACTGTTTAAATC 881

RESULT 5
BM562682 1059 bp mRNA linear EST 20-FEB-2002
LOCUS BM562682
DEFINITION AGENCOURT_6588903 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478431
5', mRNA sequence.
ACCESSION BM562682
VERSION BM562682.1 GI:18808978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1059)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

|||||
Db 937 GTRACTGTAGAGGTGTGGAGGCCCTTCATCAGGACTACATCTGAGCGCCGACATTC 878
Qy 390 TGACTGCACAGGAGAAAGCCGTTTGTGTGCAGCCAGCTGGTGCATCAAAAATTC 449
Db 877 THACTCACACAGGAGAAAGCCGTTTGTGTGC-GCCAAATGCTGTGATCAAAAATTC 819
Qy 450 ACACAAATCAAACTTGAGAAACATTTTGACGCAAAACATGAAATCAACAAAACAT 509
Db 818 ACACAAATCAAACTTGAGAAACATTTTGACGCAAAACATGAAATCAACAAAACAT 759
Qy 510 ATATATGCAAGTTTGAAGACTGTAGAAACCTTTAAGAAACATCAGACCTGAAATTC 569
Db 758 ATATATGCAAGTTTGAAGACTGTAGAAACCTTTAAGAAACATCAGACCTGAAATTC 699
Qy 570 ATGAGTCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATGTGGGAAC 629
Db 698 ATCAGTCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATGTGGGAAC 639
Qy 630 ACTTGCATCACCCAGCAGCACTGAAACGACATGCCAAGGCCCAAGGGCTATGTATGC 689
Db 638 ACTTGCATCACCCAGCAGCACTGAAACGACATGCCAAGGCCCAAGGGCTATGTATGC 579
Qy 690 AAAAGAGATGTTCTTTGTGGCAAAAACATGACGGAATCTTGAAACATGTGAGAGAAA 749
Db 578 AAAAGAGATGTTCTTTGTGGCAAAAACATGACGGAATCTTGAAACATGTGAGAGAAA 519
Qy 750 CCCATAAGAGAAATCTATGTGAATATGCCGGAACATTTAAAGCAAGATTAAC 809
Db 518 CCCATAAGAGAAATCTATGTGAATATGCCGGAACATTTAAAGCAAGATTAAC 459
Qy 810 TTAAGCAACATGMAAACTATGTCGCCAGAAAGGATGTATGCTGTCAAGAGAG 869
Db 458 TTAAGCAACATGMAAACTATGTCGCCAGAAAGGATGTATGCTGTCAAGAGAG 399
Qy 870 GCTGTGAAAGAACTATGCTGTGTATTCCTCAAGGCAATATCTCTCTCCATG 929
Db 398 GCTGTGAAAGAACTATGCTGTGTATTCCTCAAGGCAATATCTCTCTCCATG 339
Qy 930 AGGAAAGCCGCCCTTTGTGTGTAACATGCTGTGGCAAAACATTTGCAATGAAC 989
Db 338 AGGAAAGCCGCCCTTTGTGTGTAACATGCTGTGGCAAAACATTTGCAATGAAC 279
Qy 990 AAATCTCAGTAGGCAATGCTGTGTGATCATCTGACAAAGAAATGAAGCTCAAG 1049
Db 278 AAATCTCAGTAGGCAATGCTGTGTGATCATCTGACAAAGAAATGAAGCTCAAG 219
Qy 1050 TCAAAAATCTCTGAAAAAGAGGTTGGCTCTCATCTCAGTGATATATCCCTCCA 1109
Db 218 TCAAAAATCTCTGAAAAAGAGGTTGGCTCTCATCTCAGTGATATATCCCTCCA 159
Qy 1110 AAAGGAACAAGGCAAGGCTTATCTTTGTGTCAAAAAGGAGATCCCAACTGTGCG 1169
Db 158 AAAGGAACAAGGCAAGGCTTATCTTTGTGTCAAAAAGGAGATCCCAACTGTGCG 99
Qy 1170 AAGCAAGATGCTCTCGACAGTTGCACTACTTACCTTGGCTAA 1213
Db 98 AAGCAAGATGCTCTCGACAGTTGCACTACTTACCTTGGCTAA 55

RESULT 7
B0650498 930 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOUPR_8207703 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283123
DEFINITION 5', mRNA sequence.
ACCESSION B0650498
VERSION B0650498.1 GI:21774670
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 930)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCM2478 row: j column: 20
High quality sequence stop: 636.
location/Qualifiers
1. 930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6283123"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."
BASE COUNT 316 a 212 c 200 g 201 t 1 others
ORIGIN
Query Match 70.3%; Score 860.4; DB 14; Length 930;
Best Local Similarly 97.0%; Pred. No. 1.9e-180;
Matches 899; Conservative 0; Mismatches 22; Indels 6; Gaps 2;
Qy 241 CTCCTCCGTCAGTCAGCGGCATTTACAGAAAGCCTGGAAGCTTACCGGCCTGTG 300
Db 1 CTCCTCCGTCAGTCAGCGGCATTTACAGAAAGCCTGGAAGCTTACCGGCCTGTG 60
Qy 301 CAAGCAGCGGGGAGAGACCATTTGTTGTGACTATGAAGGGTGTGCAAGGCTTCAT 360
Db 61 CAAGCAGCGGGGAGAGACCATTTGTTGTGACTATGAAGGGTGTGCAAGGCTTCAT 120
Qy 361 CAGGACTACCATCTGAGCGGCCACATTTGACTCACAGAGAAAGCCGTTTGTG 420
Db 121 CAGGACTACCATCTGAGCGGCCACATTTGACTCACAGAGAAAGCCGTTTGTG 180
Qy 421 TGCAGCCAGTGGCTGTGATCAAAAATTCACACAAATCAAACTGGAAGCAATTTGA 480
Db 181 TGCAGCCAGTGGCTGTGATCAAAAATTCACACAAATCAAACTGGAAGCAATTTGA 240
Qy 481 ACGCAAAATGAAAAATCAACAAAAAATATATATGCAAGTTTGAAGCTGTAAAGAC 540
Db 241 ACGCAAAATGAAAAATCAACAAAAAATATATATGCAAGTTTGAAGCTGTAAAGAC 300
Qy 541 CTTTAAGAAACATCAGCAGCTGAAATTCATCAGTGCAGCATACCAATGACCTTATT 600
Db 301 CTTTAAGAAACATCAGCAGCTGAAATTCATCAGTGCAGCATACCAATGACCTTATT 360
Qy 601 CAATGTACCCAGGAAGGATGTGGAAACCTTTGCATCCGCCGCAAGCTGAAAGACA 660
Db 361 CAATGTACCCAGGAAGGATGTGGAAACCTTTGCATCCGCCGCAAGCTGAAAGACA 420
Qy 661 TGCAGAGCCCGCAGGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 720
Db 421 TGCAGAGCCCGCAGGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 480
Qy 721 GACGGAACCTTCTGAACATGTGACAGAAACCCATTAAGAGAGAAATCTATGTGAAGTATG 780
Db 481 GACGGAACCTTCTGAACATGTGAGAGAAACCCATTAAGAGAGAAATCTATGTGAAGTATG 540

QY 781 CCGAAAAATTTAAAGCAAGATTAACCTTAAGCAACACATGAAGAACTCATGCCCCAGA 840
|||||
Db 541 CCGAAAAATTTAAAGCAAGATTAACCTTAAGCAACACATGAAGAACTCATGCCCCAGA 600
QY 841 AAGGATGATGATGCTGCTGCTCCAGAGAGGCTGTGAAGAACCTATACTACTGTTTAA 900
|||||
Db 601 AAGGATGATGATGCTGCTGCTCCAGAGAGGCTGTGAAGAACCTATACTACTGTTTAA 660
QY 901 TCTCCAAAGCCATATCTCTCTCTCTCCATGAGAGAAAGCCGCTTTGTGTGAACATGC 960
|||||
Db 661 TCTCCAAAGCCATATCTCTCTCTCTCCATGAGAGAAAGCCGCTTTGTGTGAACATGC 720
QY 961 TGGCTGGGCAAAATTTGCAATGAAGCAAGTCACTGATGCTGTGTGAACATGA 1020
|||||
Db 721 TGGCTGGGCAAAATTTGCAATGAAGCAAGTCACTGATGCTGTGTGAACATGA 780
QY 1021 TCTGACAGAGAGAAATGAAGTCAAGTCAAGAAATCTGTAAGAAACGAGTTTGGC 1080
|||||
Db 781 TCTGACAGAGAGAAATGAAGTCAAGTCAAGAAATCTGTAAGAAACGAGTTTGGC 840
QY 1081 CTCTCATCTCTAGTGG---ATATATCCCTCCCAAGAGAAAGAGGCAAG---CTTATC 1134
|||||
Db 841 CTCTCATCTCTAGTGGATATATCCCTCCCAAGAGAAAGAGGCAAGGCTCATCTCT 900
QY 1135 TTGTGTCAAAAGGAGAGTCAACCA 1161
|||
Db 901 TTGTGTCAAAAGGAGATTCACCA 927

RESULT 8
AL527493 957 bp mRNA linear EST 13-FEB-2001
LOCUS AL527493
DEFINITION AL527493 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC024YB05 5
prime, mRNA sequence.
ACCESSION AL527493
VERSION AL527493.1 GI:12790986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS L.M.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source
1..957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC024YB05"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/issue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 283 a 240 c 227 g 206 t 1 others
ORIGIN
Query Match 70.0%; Score 849.4; DB 9; Length 957;

Best Local Similarity 91.9%; Pred. No. 5, 3e-178;
Matches 948; Conservative 1; Mismatches 7; Indels 75; Gaps 2;
QY 55 CCGGACAGTGTCTCGGACGTGACCGCGCTGCGCTTGAGAGCGCGCGC 114
|||||
Db 1 CCGGACAGTGTCTCGGACGTGACCGCGCTGCGCTTGAGAGCGCGCGC 60
QY 115 CCGTGAATCCCGGCGGTGTCGCGGAGTGGTGTCTCTTGAACATCCGACGGCTT 174
|||||
Db 61 CCGTGAATCCCGGCGGTGTCGCGGAGTGGTGTCTCTTGAACATCCGACGGCTT 120
QY 175 CATT-GGAGCGGGGAGAGCTCAGCTCCGACCGCGCGCGCGCTTCCAGAGGT 233
|||||
Db 121 CATTAGCAGCGGGGAGAGCTCAGCTCCGACCGCGCGCGCGCTTCCAGAGGT 180
QY 234 TCATCTCTCTCTCTCTGACTGACGGCCAAATTAACAGCAAGCTTGAGCGGC 293
|||||
Db 181 TCATCTCTCTCTCTCTGACTGACGGCCAAATTAACAGCAAGCTTGAGCGGC 240
QY 294 ACCGTGCAAGCAGCGGGGAGAGCAATTTGTTGTGACTATGAAGGTGGCAAG 353
|||||
Db 241 ACCGTGCAAGCAGCGGGGAGAGCAATTTGTTGTGACTATGAAGGTGGCAAG 300
QY 354 CCTTCATCAGGAGTACATCTGAGCGCCACATTTGACTCAGACAGAGAAAGCGGT 413
|||||
Db 301 CCTTCATCAGGAGTACATCTGAGCGCCACATTTGACTCAGACAGAGAAAGCGGT 360
QY 414 TTGTTGTGACAGCCAGTGTGTGATCAAAATTAACAGCAAGCTTGAGAAAC 473
|||||
Db 361 TTGTTGTGACAGCCAGTGTGTGATCAAAATTAACAGCAAGCTTGAGAAAC 420
QY 474 ATTTTGAACGCAATCAATGAATCAACAAATATATATGAGTTTGAACAGCTGA 533
|||||
Db 421 ATTTTGAACGCAATCAATGAATCAACAAATATATATGAGTTTGAACAGCTGA 480
QY 534 AGAAGACCTTTAAGAAATCAGCAGCTGAAATTCATGAGTGCACATACCAATGAAC 593
|||||
Db 481 AGAAGACCTTTAAGAAATCAGCAGCTGAAATTCATGAGTGCACATACCAATGAAC 540
QY 594 CTCTATTCAAGTGTACCCAGAGAGATGTGGAAACATTTGATCAACAGCAAGCTGA 653
|||||
Db 541 CTCTATTCAAGTGTACCCAGAGAGATGTGGAAACATTTGATCAACAGCAAGCTGA 550
QY 654 AAGCAGATGCCAAGGCCACAGGCTATGTATGTCAAAAGATGTTCCTTTGTGCAA 713
|||||
Db 551 -----GCTATGTATGTCAAAAGATGTTCCTTTGTGCAA 586
QY 714 AAACATGAGCGAATCTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTG 773
|||||
Db 587 AAACATGAGCGAATCTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTG 646
QY 774 AAGTATGCCGGAAGAAATTTAAAGCGAAAGATTAACCTTAAGCAACATGAAGCTATG 833
|||||
Db 647 AAGTATGCCGGAAGAAATTTAAAGCGAAAGATTAACCTTAAGCAACATGAAGCTATG 706
QY 834 CCCGAGAAAGGAGTATGTGCTGTCCAGAGAAAGCTGTGAGAAACCTATACTACTG 893
|||||
Db 707 CCCGAGAAAGGAGTATGTGCTGTCCAGAGAAAGCTGTGAGAAACCTATACTACTG 766
QY 894 TGTTTAATCTCAAGCATATCTCTCTCTCTCTGATGAGAGAAAGCGCGCTTTGTGTG 953
|||||
Db 767 TGTTTAATCTCAAGCATATCTCTCTCTCTCTGATGAGAGAAAGCGCGCTTTGTGTG 826
QY 954 AACATGCTGCTGTGGCAAAACATTTCCATGAAGAAACAAAGTCTCAGTGGCATCTGTG 1013
|||||
Db 827 AACATGCTGCTGTGGCAAAACATTTCCATGAAGAAACAAAGTCTCAGTGGCATCTGTG 886
QY 1014 TACATGATCTCGAAGAGAAATGAAGCTCAAGATCAAAATTAATTCGTGAGAAAGCGA 1073
|||||
Db 887 TACATGATCTCGAAGAGAAATGAAGCTCAAGATCAAAATTAATTCGTGAGAAAGCGG 946
QY 1074 GTTTGGCCTCT 1084
|||

BASE COUNT	250 a	232 c	230 g	177 t	4 others
ORIGIN					
Query Match	69.2%: Score 838.8; DB 9; Length 893;				
Best Local Similarity	99.4%: Pred. No. 1.2e-175;				
Matches 840; Conservative	0; Mismatches 5; Indels 0; Gaps 0;				
QY	1	GTGCGGCGCGCGGAGGTTTCAGACAGGAGCGGTGGCGCGCGCGCGGTCCCGGC	60		
Db	2	GTGCGGCGGTGGCGGAGGTTTCAGACAGGAGCGGTGGCGCGCGCGCGGTCCCGGC	61		
QY	61	ACGTGTCTGCGACGTGGCAGCGCGCTGGCCCTTGAGAGCGCGCGCGCTTGA	120		
Db	62	ACGTGTCTGCGACGTGGCAGCGCGCTGGCCCTTGAGAGCGCGCGCGCTTGA	121		
QY	121	TCCGCGCGCGCGTGGTCCCGGAGTGGGTGCTCTTACACATCCGCGCGCTTATTC	180		
Db	122	TCCGCGCGCGCGTGGTCCCGGAGTGGGTGCTCTTACACATCCGCGCGCTTATTC	181		
QY	181	AGCGGCGAGAGCTCAGCTCGACCGCGCGCGCGCGCTTCCAGAGGTTTCATCTG	240		
Db	182	AGCGGCGAGAGCTCAGCTCGACCGCGCGCGCGCGCTTCCAGAGGTTTCATCTG	241		
QY	241	CTCCTTCCCTGACTGACAGCGCCCAATTACAGCAAGCCTGGAAGCTTGACGCGCACCTGTG	300		
Db	242	CTCCTTCCCTGACTGACAGCGCCCAATTACAGCAAGCCTGGAAGCTTGACGCGCACCTGTG	301		
QY	301	CAACACACAGCGGAGAGACCATTTGTTGACTATTAAGAGGTGTGGCAAGGCTTAT	360		
Db	302	CAACACACAGCGGAGAGACCATTTGTTGACTATTAAGAGGTGTGGCAAGGCTTAT	361		
QY	361	CAGGAGACTACATGTGACCGCGCACATCTGACTCACAGAGAGAAAGCGTTGTTG	420		
Db	362	CAGGAGACTACATGTGACCGCGCACATCTGACTCACAGAGAGAAAGCGTTGTTG	421		
QY	421	TGCAGCCACTGGCTGTGATCAAAATTCACACAAATCAAACTGAAGAAACATTTTGA	480		
Db	422	TGCAGCCACTGGCTGTGATCAAAATTCACACAAATCAAACTGAAGAAACATTTTGA	481		
QY	481	ACGCAAAATGAAAAATCAACAAAAACATATATATGAGTTTGAAGCTTGAAGAC	540		
Db	482	ACGCAAAATGAAAAATCAACAAAAACATATATATGAGTTTGAAGCTTGAAGAC	541		
QY	541	CTTTAAGAATCATGACAGCTGAAAAATCATCATGAGCATACCAATGAACCTCTATT	600		
Db	542	CTTTAAGAATCATGACAGCTGAAAAATCATCATGAGCATACCAATGAACCTCTATT	601		
QY	601	CAAGTGTACCCAGAGAGGATGTGGAAAACTTTGCATCACCCAGCAAGTGAACGACA	660		
Db	602	CAAGTGTACCCAGAGAGGATGTGGAAAACTTTGCATCACCCAGCAAGTGAACGACA	661		
QY	661	TGCCAAGGCCACAGAGGCTATGTATGCAAAAAGGATGTTCTTTTGCGCAAAAACATG	720		
Db	662	TGCCAAGGCCACAGAGGCTATGTATGCAAAAAGGATGTTCTTTTGCGCAAAAACATG	721		
QY	721	GACGGAATCTTGAACATGTGAGAGAAACCATTAAGAGAAATACTATGGAAGTATG	780		
Db	722	GACGGAATCTTGAACATGTGAGAGAAACCATTAAGAGAAATACTATGGAAGTATG	781		
QY	781	CCGGAATAATTTAAACGCAAGATTTACCTTAAGCAACATGAATACTCATGCCCCANA	840		
Db	782	CCGGAATAATTTAAACGCAAGATTTACCTTAAGCAACATGAATACTCATGCCCCANA	841		
QY	841	AAGG 845			
Db	842	AANG 846			

RESULT 11
 LOCUS B0644275 902 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8355851 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285421
 5', mRNA sequence.

ACCESSION	B0644275
VERSION	B0644275.1
KEYWORDS	GI:21768447
SOURCE	ESF.
ORGANISM	human.
	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 902)
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs@remail.nih.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LCM2484 row: j column: 14 High quality sequence stop: 642.
FEATURES	Location/Qualifiers
source	1. 902 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6285421" /clone_lib="NIH_MGC_100" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: liver; Vector: pORF8; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGACGAG(5). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."
BASE COUNT	301 a 208 c 197 g 194 t 2 others
ORIGIN	
Query Match	68.9%; Score 836; DB 14; Length 902;
Best Local Similarity	98.3%; Pred. No. 4.9e-175;
Matches 875; Conservative	0; Mismatches 12; Indels 3; Gaps 3;
QY	240 GCTCCTTCCCTGACTGACGCGCAATTAACAGCAAGCCTGGAAGCTTGACGCGACCTGT 299
Db	11 GCTCCTTCCCTGACTGACGCGCAATTAACAGCAAGCCTGGAAGCTTGACGCGACCTGT 70
QY	300 GCAAGCACAGGAGGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGGCTTCA 359
Db	71 GC-ANACACAGGAGGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGGCTTCA 129
QY	360 TCAGGAGTACACTGTGAGCGCGCACATTTGACTGCACAGAGAGAAAGCGTTTGT 419
Db	130 TCAGGAGTACACTGTGAGCGCGCACATTTGACTGCACAGAGAGAAAGCGTTTGT 189
QY	420 GTGCAAGCCACTGGCTGTGATCAAAATTCACACAAATCAAACTGAAGAAATTTTG 479
Db	190 GTGCAAGCCACTGGCTGTGATCAAAATTCACACAAATCAAACTGAAGAAATTTTG 249
QY	480 AACGCAATGAAATTAACAAAAATATATATGACGTTTGAAGACGTGAAGAAGA 539
Db	250 AACGCAATGAAATTAACAAAAATATATATGACGTTTGAAGACGTGAAGAAGA 309
QY	540 CCTTTAAGAATCATGACAGCTGAAATTCATGAGTGCACATACCAATGAACCTCTAT 599
Db	310 CCTTTAAGAATCATGACAGCTGAAATTCATGAGTGCACATACCAATGAACCTCTAT 369
QY	600 TCAAGTGTACCCAGAGAGATGTGGAAACACTTTGATCAACGCAAGCTGAAGAC 659
Db	370 TCAAGTGTACCCAGAGAGATGTGGAAACACTTTGATCAACGCAAGCTGAAGAC 429

QY 660 ATGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACAT 719
|||||
Db 430 ATGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACAT 489
|||||
QY 720 GGACGGAACCTTGTGAACATGTGTGAGAGAAACCATTAAGAGGAATATCTATGTGAAGTAT 779
|||||
Db 490 GGACGGAACCTTGTGAACATGTGTGAGAGAAACCATTAAGAGGAATATCTATGTGAAGTAT 549
|||||
QY 780 GCCGGAACCATTTAAACGCAAGATTACCTTAAGCAACATGAACATCATGCCCCAG 839
|||||
Db 550 GCCGGAACCATTTAAACGCAAGATTACCTTAAGCAACATGAACATCATGCCCCAG 609
|||||
QY 840 AAGAGGATGATGTGCTGTGTCCAGAGAGGCTGTGAGAGAACCTTAATCACTGTGTTTA 899
|||||
Db 610 AAGAGGATGATGTGCTGTGTCCAGAGAGGCTGTGAGAGAACCTTAATCACTGTGTTTA 669
|||||
QY 900 ATCTCCAAAGCCATATCTCTCTCCCTTCATGAGAGAACGCCCTTTTGTGTGTGAACATG 959
|||||
Db 670 ATCTCCAAAGCCATATCTCTCTCCCTTCATGAGAGAACGCCCTTTTGTGTGTGAACATG 729
|||||
QY 960 CTGCGCTGGCAAAACATTTGCAATGAACAAAGCTCACTAGGCAATGCTGTGTACATG 1019
|||||
Db 730 CTGCGCTGGCAAAACATTTGCAATGAACAAAGCTCACTAGGCAATGCTGTGTACATG 789
|||||
QY 1020 ATCTGCAAGAGAAATGATGAAGCTCAAAAGTCAAAAATCTCTGTAAGAAACGAGTTTG 1079
|||||
Db 790 ATCTGCAAGAGAAATGATGAAGCTCAAAAGTCAAAAATCTCTGTAAGAAACGAGTTTG 848
|||||
QY 1080 CCTCTCATCTCATGTGATATAT - CCTCTCCCAAAAGAGAAACGAGGCAAG 1128
|||||
Db 849 CCTCTCATCTCATGTGATATAT - CCTCTCCCAAAAGAGAAACGAGGCAAG 898
|||||

RESULT 12
AL578193/c 933 bp mRNA linear EST 16-FEB-2001
LOCUS AL578193 LTI_NFL006_PU2 Homo sapiens cDNA clone CS0DK002YJ18 3
DEFINITION prime, mRNA sequence.
ACCESSION AL578193
VERSION AL578193.1 GI:12942034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS L.W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1. 933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YJ18"
/clone_1lb="LTI_NFL006_PU2"
/lssue_type="placenta"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 202 a 192 c 207 g 326 t 6 others
ORIGIN

Query Match 68.4%; Score 830.2; DB 9; Length 933;
Best Local Similarity 98.5%; Pred. No. 9,5e-174;
Matches 854; Conservative 4; Mismatches 7; Indels 2; Gaps 2;

QY 348 GCAAGCCCTTCATCAGGAGCTACCATGTGAGCGCCCATCTTGACTGCACAGAGAAA 407
|||||
Db 933 GCAAGCCCTTCATCA - GGACTACCATGTGAGCGCCCATCTTGACTGCACAGAGAAA 875
|||||
QY 408 AGCCGTTGTTGTGCGAGCCGCTGTGTGATCAAAAATTCACACAAAATCAACTTGA 467
|||||
Db 874 AGCCGTTGTTGTGCGAGCCGCTGTGTGATCAAAAATTCACACAAAATCAACTTGA 815
|||||
QY 468 AGAAGATTTTGAAGCGCAACATGAATAACAAACAAATATATATGTGAGTTTGAAG 527
|||||
Db 814 AGAAGATTTTGAAGCGCAACATGAATAACAAACAAATATATATGTGAGTTTGAAG 755
|||||
QY 528 ACTGTAGAGACCTTTAAGAACATGACAGCTGAAAATTCATCACTGCCAGATACCA 587
|||||
Db 754 ACTGTAGAGACCTTTAAGAACATGACAGCTGAAAATTCATCACTGCCAGATACCA 695
|||||
QY 588 ATGAACCTTATTCAGGTACCGAGAGAGATGTGGAAACACTTGTGCATCACCGACA 647
|||||
Db 694 ATGAACCTTATTCAGGTACCGAGAGAGATGTGGAAACACTTGTGCATCACCGACA 635
|||||
QY 648 AGCTGAAGCGACATGCCAAGGCCAGAGGCTATGTATGTCAAAAAGATGTTCTTTG 707
|||||
Db 634 AGCTGAAGCGACATGCCAAGGCCAGAGGCTATGTATGTCAAAAAGATGTTCTTTG 575
|||||
QY 708 TGGCAAAACATGAGAGGAACTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATAC 767
|||||
Db 574 TGGCAAAACATGAGAGGAACTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATAC 515
|||||
QY 768 TATGTGAAGTATGCGGGAACATTTAAAGCAAAAGTACCTTAACCAACACATGAAA 827
|||||
Db 514 TATGTGAAGTATGCGGGAACATTTAAAGCAAAAGTATTCCTTAACGCACACATGAAA 455
|||||
QY 828 CTCATGCCCCGAAAGGAGATGATGTGCTGTCCAGAGAGAGCTGTGAGAGAACTATA 887
|||||
Db 454 CTCATGCCCCGAAAGGAGATGATGTGCTGTCCAGAGAGAGCTGTGAGAGAACTATA 395
|||||
QY 888 CTACTGTGTTTATCTCCAAAGCATATCTCTCTTCATGAGAGAACGCCCTTTTG 947
|||||
Db 394 CAACCTGTGTTTATCTCCAAAGCATATCTCTCTTCATGAGAGAACGCCCTTTTG 335
|||||
QY 948 TGTGTGAACATGCGGCTGTGGCAAAACATTTGCAATGAAGCAAAAGTCTCACTAGGCAATG 1007
|||||
Db 334 TGTGTGAACATGCGGCTGTGGCAAAACATTTGCAATGAAGCAAAAGTCTCACTAGGCAATG 275
|||||
QY 1008 CTGTTGTACATGATCTGACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGTGAAA 1067
|||||
Db 274 CTGTTGTACATGATCTGACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGTGAAA 215
|||||
QY 1068 AACGAGATTTGGCTCTCATCTCAGTGTATATTCCTCCCAAAAGAGAGGCAAG 1127
|||||
Db 214 AACGAGATTTGGCTCTCATCTCAGTGTATATTCCTCCCAAAAGAGAGGCAAG 155
|||||
QY 1128 GCTATCTTTGTGCAAAAGCGAGATCACCAACTGTGTGGAAGACAGATGTCTGCA 1187
|||||
Db 154 GCTATCTTTGTGCAAAAGCGAGATCACCAACTGTGTGGAAGACAGATGTCTGCA 95
|||||
QY 1188 CAGTTGCA - GTACTTACCTTGGCTAA 1213
|||||
Db 94 CAGTTGCAAGCAGCTTACCTTGGCTAA 68
|||||

RESULT 13
BO278693 1089 bp mRNA linear EST 07-MAY-2002
LOCUS BO278693
DEFINITION AGENCOURT_7049694 NIH_MGC_107 Homo sapiens cDNA clone IMAGE:5805333
5', mRNA sequence.
ACCESSION BO278693
VERSION BO278693.1 GI:20488901

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1089)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LICM2044 row: f column: 22 High quality sequence stop: 610.
FEATURES	
source	1. .1089 Location/Qualifiers /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5805333" /clone_lib="NIH_MGC_107" /tissue_type="adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: breast; Vector: pORF7; Site:1: EcoRI; Site:2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
BASE COUNT	296 a 289 c 283 g 221 t
ORIGIN	
Query Match	68.3%; Score 829; DB 14; Length 1089;
Best Local Similarity	94.6%; Pred. No. 1.8e-173;
Matches 926: Conservative	0; Mismatches 40; Indels 13; Gaps 6;
11	CGCGCGAAGGTTTCAGCAGGAGCGGTGGGCGCGCGCGCGGTCCCGGCAGCGTGTGCG 70
2	CGCGCGAAGGTTTCAGCAGGAGCGGTGGGCGCGCGCGCGGTCCCGGCAGCGTGTGCG 61
71	GCACGTGGCAGCGGCGCCCTGCGCTTGGGCTTGGAGGCGCGCGCGCGCGTGTGCGCGCC 130
62	GCACGTGGCAGCGGCGCGCCCTGCGCTTGGAGGCGCGCGCGCGCGTGTGCGCGCGCC 121
131	GTGGTCGCGCGAATCGGTGCGTCTTGACACATCGGCGAGCGGTTCATTGACGCGGCGAG 190
122	GTGGTCGCGCGAATCGGTGCGTCTTGACACATCGGCGAGCGGTTCATTGACGCGGCGAG 181
191	AGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCCAGGAGGTTATCGTCTCTTCCT 250
182	AGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCCAGGAGGTTATCGTCTCTTCCT 241
251	GACCTCAGCGCCCAATTACAGCAAACCGTTGGAAGCTTGGACGCGCAGCTGTGCAAGCACAG 310
242	GACCTCAGCGCCCAATTACAGCAAACCGTTGGAAGCTTGGACGCGCAGCTGTGCAAGCACAG 301
311	GGGAGAGAGCCATTGTTGTTGACTATGAAAGGCTGTGGCAAGGCGCTTCATCAGGGAGCTAC 370
302	GGGAGAGAGCCATTGTTGTTGACTATGAAAGGCTGTGGCAAGGCGCTTCATCAGGGAGCTAC 361
371	CATCTGAGCGCGCACATCTTGACTCACACAGAGAGAAAAGCCGTTTGTGTTGTGCAGCCACT 430
362	CATCTGAGCGCGCACATCTTGACTCACACAGAGAGAAAAGCCGTTTGTGTTGTGCAGCCACT 421
431	GGCTGTGATCAAAATTCACACAAAATCAAACTTGAGAGAAACATTTTGAACGCAACAT 490

Db	422	GCGTGTGATCAAAATTTCACAGCAAATTAACAACCTTGGAAGAAAACATTTTGGAACGCCAATC	481
OY	491	GAATAATCAACAAAAACAATATATATATGACGATTTTGAAGAAGCTGTAAAGAACCTTTAAGAAA	550
Db	482	GAATAATCAACAAAAACAATATATATGACGATTTTGAAGAAGCTGTAAAGAACCTTTAAGAAA	541
OY	551	CATCAGCACGCTGAAAATCCATCATGAGCCACAGCATACCATTGAACCTCTATTTCAGTGTACC	610
Db	542	CATCAGCACGCTGAAAATCCATCATGAGCCACAGCATACCATTGAACCTCTATTTCAGTGTACC	601
OY	611	CAGSAGAGTGTGGGAAAACATTTGCATCCACACCAAGCTGAAAAGCAGATGCCAAGGCC	670
Db	602	CAGSAGAGTGTGGGAAAACATTTGCATCCACACCAAGCTGAAAAGCAGATGCCAAGGCC	661
OY	671	CACGAGGCGTAGTATGTATCTCAAAAAAGAGATGTCCTTTGTGGCAAAAACATGACGCAACTT	730
Db	662	CACGAGGCGTAGTATGTATCTCAAAAAAGAGATGTCCTTTGTGGCAAAAACATGACGCAACTT	721
OY	731	CTGAACACATGTGTAGAGAAAACCCATTAAAGAGAAATCTATGTGAAGTATGCCGGAAAAACA	790
Db	722	CTGAACACATGTGTAGAGAAAACCCATTAAAGAGAAATCTATGTGTGAAGTATGCCGGAAAAACA	781
OY	791	TPTTAAGCCAAGATATTACCTTAAAGCAACATGAAACTCAAT - GCCCCAGAAAGGGATG	848
Db	782	TPTTAAGCCAAGATATTACCTTAAAGCAACATGAAACTCAAT - GCCCCAGAAAGGGATG	841
OY	849	TATGTGCG - TGTCACAGAGAAGCGTGTGG - AAGAACCTATTACTACTGT - GTTTAATCT	903
Db	842	TATGTGCGCTGTGTGCCAAAAGAAAGTGTGTGGA AAAAACCTATTCCACTGTGGTTTAATCT	901
OY	904	CCAAAGCCATATC -- CTCTCCCTTCATGAGGAAAGCGCCCTTTTT -- GTGTGTGAACA	957
Db	902	CCCAAGCCATATCCTCTCTCTCTTCCCTGAAGAAAGCGCCCTTTTGGGGGGGGAACA	961
OY	958	TGCTGCTGTGTGGCAAAACA	976
Db	962	TGCTGCTGTGTGGCGCAAAA	980
RESULT 14 BG830080	BG830080	927 bp mRNA linear EST 22-MAY-2001	
LOCUS	602764826E1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4906710 5'		
DEFINITION	mRNA sequence.		
ACCESSION	BG830080		
VERSION	BG830080.1 GI:14177667		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 927) NIH-MGC http://mgc.ncl.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC)		
AUTHORS	Unpublished (1999)		
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: csapbs@mail.nih.gov		
COMMENT	Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN). DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LNCMI810 row: d column: 07 High quality sequence stop: 880. Location/Qualifiers		
FEATURES	source		
	1..927	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:4906710"	
		/clone_lib="NIH_MGC_42"	
		/tissue_type="epithelioid carcinoma cell line"	


```
Db 181 TGCAGCCAAATGGCTGTGATCAAAAAATTCACACAAATCAAACTTGAAGAAACATTTGA 240
QY 481 ACGGAACATGAAATATCAACAAAAACATATATATGCACTTTGAAGACTGTAGAAGAC 540
Db 241 ACGCAAAACATGAAAAATCAACAAAAACAATATATATGCGATTTGAGACTGTAGAAGAC 300
QY 541 CTTTAAGAACATGACGAGCTGAAAAATCCATCACTGCCAGCATACCAATGAACCTTATT 600
Db 301 CTTTAAGAACATGACGAGCTGAAAAATCCATCACTGCCAGCATACCAATGAACCTTATT 360
QY 601 CAAGTGTACCCAGAGAGATGTGGAAAACTTTGCATCACCCAGCAAGCTGAACGACA 660
Db 361 CAAGTGTACCCAGAGAGATGTGGAAAACTTTGCATCACCCAGCAAGCTGAACGACA 420
QY 661 TGCAGAGCCCGACGAGGCTATGTATGTCAAAAAAGATGTTCTTTGTGGCAAAAACATG 720
Db 421 TGCCAGGCCCCACGAGGCTATGTATGTCAAAAAAGATGTTCTTTGTGGCAAAAACATG 480
QY 721 GACGGAACCTTCTGAACATGTGAGAGAACCCATTAAGAGAAATACTATGTGAATATG 780
Db 481 GACGGAACCTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAATATG 540
QY 781 CCGGAAAAACATTTAAGCGCAAGATTAACCTTAACCAACACATGAACCTCATGCCCGAGA 840
Db 541 CCGGAAAAACATTTAAGCGCAAGATTAACCTTAACCAACACATGAACCTCATGCCCGAGA 600
QY 841 AAGGATGTATGTGCTGTCCAAAGAGGCTGTGGAAGAACTATACTGTGTTAA 900
Db 601 AAGGATGTATGTGCTGTCCAAAGAGGCTGTGGAAGAACTATACTGTGTTAA 660
QY 901 TCTCCAAAGCCATATCTCTCTCCATGAGAG--AAAGCCGCCCTTTGTGTGTAACAT 958
Db 661 TCTCCAAAGCCATATCTCTCTCTCCATGAGGAAAGCCGCCCTTTGTGTGTAACAT 720
QY 959 GCTGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCACTAGCAATGCTGTGTACAT 1018
Db 721 GCTGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCCTAGCAATGCTGTGTACAT 780
QY 1019 GATTCCTGACAAAGAAATGAAGCTCAAAAGTCAAAAAATCTGTAAGAAACGAGTTTG 1078
Db 781 GATTCCTGACAAAG--AGAAATGAAGCTCAAAAGTCAAAAAATCTGTAAGAAACGAGTTTG 839
QY 1079 G-CCTTCATCTCAGTGAATATCC--TCCCAAAAGGAAACAAAGGCAAGGCTTATCT 1135
Db 840 GCCCTTCATCTCAGTGAATATCCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCT 899
QY 1136 TTGTGTCA 1143
Db 900 TTGGGGCA 907
```

Search completed: February 10, 2003, 17:44:18
Job time : 1907.59 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:45:57 ; Search time 40 Seconds
(without alignments)
739.844 Million cell updates/sec

Title: US-09-831-426C-2
Perfect score: 2008
Sequence: 1 MDPVAVAESVSLTIADAF.....ESPNCVEDKMLSTVALTLG 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 405691 segs, 81078759 residues

Total number of hits satisfying chosen parameters: 405691

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2000	99.6	430	US-09-724-676-93212	Sequence 93212, A
2	2000	99.6	430	US-09-724-676-93223	Sequence 93223, A
3	2000	99.6	430	US-09-724-676-93234	Sequence 93234, A
4	2000	99.6	430	US-09-724-676-93245	Sequence 93245, A
5	2000	99.6	430	US-09-724-676-93256	Sequence 93256, A
6	2000	99.6	430	US-09-724-676-93267	Sequence 93267, A
7	2000	99.6	430	US-09-724-676-93278	Sequence 93278, A
8	2000	99.6	430	US-09-724-676-93289	Sequence 93289, A
9	2000	99.6	430	US-09-724-676-93300	Sequence 93300, A
10	2000	99.6	430	US-09-724-676A-93212	Sequence 93212, A
11	2000	99.6	430	US-09-724-676A-93223	Sequence 93223, A
12	2000	99.6	430	US-09-724-676A-93234	Sequence 93234, A
13	2000	99.6	430	US-09-724-676A-93245	Sequence 93245, A
14	2000	99.6	430	US-09-724-676A-93256	Sequence 93256, A
15	2000	99.6	430	US-09-724-676A-93267	Sequence 93267, A
16	2000	99.6	430	US-09-724-676A-93278	Sequence 93278, A
17	2000	99.6	430	US-09-724-676A-93289	Sequence 93289, A
18	2000	99.6	430	US-09-724-676A-93300	Sequence 93300, A
19	1752	87.3	391	US-09-724-676-93213	Sequence 93213, A
20	1752	87.3	391	US-09-724-676-93214	Sequence 93214, A
21	1752	87.3	391	US-09-724-676-93215	Sequence 93215, A
22	1752	87.3	391	US-09-724-676-93216	Sequence 93216, A
23	1752	87.3	391	US-09-724-676-93217	Sequence 93217, A
24	1752	87.3	391	US-09-724-676-93218	Sequence 93218, A
25	1752	87.3	391	US-09-724-676-93219	Sequence 93219, A
26	1752	87.3	391	US-09-724-676-93220	Sequence 93220, A

27	1752	87.3	391	US-09-724-676-93221	Sequence 93221, A
28	1752	87.3	391	US-09-724-676A-93213	Sequence 93213, A
29	1752	87.3	391	US-09-724-676A-93214	Sequence 93214, A
30	1752	87.3	391	US-09-724-676A-93215	Sequence 93215, A
31	1752	87.3	391	US-09-724-676A-93216	Sequence 93216, A
32	1752	87.3	391	US-09-724-676A-93217	Sequence 93217, A
33	1752	87.3	391	US-09-724-676A-93218	Sequence 93218, A
34	1752	87.3	391	US-09-724-676A-93219	Sequence 93219, A
35	1752	87.3	391	US-09-724-676A-93220	Sequence 93220, A
36	1752	87.3	391	US-09-724-676A-93221	Sequence 93221, A
37	1595	79.4	354	US-09-724-676-93222	Sequence 93222, A
38	1595	79.4	354	US-09-724-676-93223	Sequence 93223, A
39	1595	79.4	354	US-09-724-676-93224	Sequence 93224, A
40	1595	79.4	354	US-09-724-676-93225	Sequence 93225, A
41	1595	79.4	354	US-09-724-676-93226	Sequence 93226, A
42	1595	79.4	354	US-09-724-676-93227	Sequence 93227, A
43	1595	79.4	354	US-09-724-676-93228	Sequence 93228, A
44	1595	79.4	354	US-09-724-676-93229	Sequence 93229, A
45	1595	79.4	354	US-09-724-676-93230	Sequence 93230, A
				US-09-724-676-93231	Sequence 93231, A

ALIGNMENTS

RESULT 1
US-09-724-676-93212
; Sequence 93212, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: Patent version 3.2
; SEQ ID NO 93212
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-93212

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 46-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MDPVAVAESVSLTIADAFIAAGESSAPPPRALPRRFICSPDCCSANYSKAMRLDAH	60
DB	66	LDPVAVAESVSLTIADAFIAAGESSAPPPRALPRRFICSPDCCSANYSKAMRLDAH	125
QY	61	ICKHTGERPPVCDYEGCGAFIRIDYLSRIILTHTEKPEVCNATGCDOKFNTKSNLKKH	120
DB	126	LCKHTGERPPVCDYEGCGAFIRIDYLSRIILTHTEKPEVCNATGCDOKFNTKSNLKKH	185
QY	121	FERKHENNOQYVICSPEDECKKTFKHHQOLKHOCQHTNEPLFKCTOBGCGKHPASPSTLK	180
DB	186	FERKHENNOQYVICSPEDECKKTFKHHQOLKHOCQHTNEPLFKCTOBGCGKHPASPSTLK	245
QY	181	RAKKAHEGYVCOGCSFVAKTWTLLKXVRETHKEILCEVCKTKRKDYLKOHKHTTA	240
DB	246	RAKKAHEGYVCOGCSFVAKTWTLLKXVRETHKEILCEVCKTKRKDYLKOHKHTTA	305
QY	241	PERDVRCRPEGCGRTYTFVFNLSHILSFHESRPFVCHACGKTFAMKOSTLTHNAV	300
DB	306	PERDVRCRPEGCGRTYTFVFNLSHILSFHESRPFVCHACGKTFAMKOSTLTHNAV	365
QY	301	HDDPKKMKLTKVKKSEKRSLSHLSGYIPPKRKQCGSLCNGESPNCVEDKMLSTVA	360
DB	366	HDDPKKMKLTKVKKSEKRSLSHLSGYIPPKRKQCGSLCNGESPNCVEDKMLSTVA	425
QY	361	VLTLG 365	
DB	426	VLTLG 430	

```
RESULT 2
US-09-724-676-93223
: Sequence 93223, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 93223
: LENGTH: 430
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-724-676-93223

Query Match          99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPRAVAESVSLTIADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 60
Db :|||||
66 LDPRAVAESVSLTIADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 125
QY 61 LCKHNGERPFVCDYEGCGKAFIRDYHLSRHILTHNGEKPFVCAATGCDOKFNTKSNLKKH 120
Db 126 LCKHNGERPFVCDYEGCGKAFIRDYHLSRHILTHNGEKPFVCAANCDDOKFNTKSNLKKH 185
QY 121 FEKKHNOOKOYICSFEDCKKTFKKHQOLKIHCQHTNBPFLKCTOEGCGKHPASPSKLK 180
Db 186 FEKKHNOOKOYICSFEDCKKTFKKHQOLKIHCQHTNBPFLKCTOEGCGKHPASPSKLK 245
QY 181 RHAKAHEGYVCOGCSFVAKTWTLLKHVRETHKEELICEVCRKTFKRKYDLKQHKHTHA 240
Db 246 RHAKAHEGYVCOGCSFVAKTWTLLKHVRETHKEELICEVCRKTFKRKYDLKQHKHTHA 305
QY 241 PERDVCRCPREGCGRTYTTVFNLQSHILSFHESRPFVCEHAGCGCTFAMKOSLTRHAVY 300
Db 306 PERDVCRCPREGCGRTYTTVFNLQSHILSFHESRPFVCEHAGCGCTFAMKOSLTRHAVY 365
QY 301 HDPDKKKMLKYKSKREKSLASHLSGYIPPKKOGGSLCONGESPPNCVEDKMLSTVA 360
Db 366 HDPDKKKMLKYKSKREKSLASHLSGYIPPKKOGGSLCONGESPPNCVEDKMLSTVA 425
QY 361 VLTIG 365
Db 426 VLTIG 430

RESULT 3
US-09-724-676-93234
: Sequence 93234, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 93234
: LENGTH: 430
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-724-676-93234

Query Match          99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPRAVAESVSLTIADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 60
```

```
Db :|||||
66 LDPRAVAESVSLTIADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 125
QY 61 LCKHNGERPFVCDYEGCGKAFIRDYHLSRHILTHNGEKPFVCAATGCDOKFNTKSNLKKH 120
Db 126 LCKHNGERPFVCDYEGCGKAFIRDYHLSRHILTHNGEKPFVCAANCDDOKFNTKSNLKKH 185
QY 121 FEKKHNOOKOYICSFEDCKKTFKKHQOLKIHCQHTNBPFLKCTOEGCGKHPASPSKLK 180
Db 186 FEKKHNOOKOYICSFEDCKKTFKKHQOLKIHCQHTNBPFLKCTOEGCGKHPASPSKLK 245
QY 181 RHAKAHEGYVCOGCSFVAKTWTLLKHVRETHKEELICEVCRKTFKRKYDLKQHKHTHA 240
Db 246 RHAKAHEGYVCOGCSFVAKTWTLLKHVRETHKEELICEVCRKTFKRKYDLKQHKHTHA 305
QY 241 PERDVCRCPREGCGRTYTTVFNLQSHILSFHESRPFVCEHAGCGCTFAMKOSLTRHAVY 300
Db 306 PERDVCRCPREGCGRTYTTVFNLQSHILSFHESRPFVCEHAGCGCTFAMKOSLTRHAVY 365
QY 301 HDPDKKKMLKYKSKREKSLASHLSGYIPPKKOGGSLCONGESPPNCVEDKMLSTVA 360
Db 366 HDPDKKKMLKYKSKREKSLASHLSGYIPPKKOGGSLCONGESPPNCVEDKMLSTVA 425

RESULT 4
US-09-724-676-93245
: Sequence 93245, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 93245
: LENGTH: 430
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-724-676-93245

Query Match          99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPRAVAESVSLTIADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 60
Db :|||||
66 LDPRAVAESVSLTIADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 125
QY 61 LCKHNGERPFVCDYEGCGKAFIRDYHLSRHILTHNGEKPFVCAATGCDOKFNTKSNLKKH 120
Db 126 LCKHNGERPFVCDYEGCGKAFIRDYHLSRHILTHNGEKPFVCAANCDDOKFNTKSNLKKH 185
QY 121 FEKKHNOOKOYICSFEDCKKTFKKHQOLKIHCQHTNBPFLKCTOEGCGKHPASPSKLK 180
Db 186 FEKKHNOOKOYICSFEDCKKTFKKHQOLKIHCQHTNBPFLKCTOEGCGKHPASPSKLK 245
QY 181 RHAKAHEGYVCOGCSFVAKTWTLLKHVRETHKEELICEVCRKTFKRKYDLKQHKHTHA 240
Db 246 RHAKAHEGYVCOGCSFVAKTWTLLKHVRETHKEELICEVCRKTFKRKYDLKQHKHTHA 305
QY 241 PERDVCRCPREGCGRTYTTVFNLQSHILSFHESRPFVCEHAGCGCTFAMKOSLTRHAVY 300
Db 306 PERDVCRCPREGCGRTYTTVFNLQSHILSFHESRPFVCEHAGCGCTFAMKOSLTRHAVY 365
QY 301 HDPDKKKMLKYKSKREKSLASHLSGYIPPKKOGGSLCONGESPPNCVEDKMLSTVA 360
Db 366 HDPDKKKMLKYKSKREKSLASHLSGYIPPKKOGGSLCONGESPPNCVEDKMLSTVA 425
```

OY 361 VLTG 365
|||||
Db 426 VLTG 430

RESULT 5
US-09-724-676-93256

; Sequence 93256, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93256
; LENGTH: 430
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-93256

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPYVAESVSLTTADAFIAGESSAPPPRPPALPRRFICFPDCCSANYSKAMKLDH 60
:|||||
Db 66 LDPYVAESVSLTTADAFIAGESSAPPPRPPALPRRFICFPDCCSANYSKAMKLDH 125
OY 61 LCKHTEGEPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
:|||||
Db 126 LCKHTEGEPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 185
OY 121 FERKHNQOKOYICSPEDCKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHPASPSK 180
:|||||
Db 186 FERKHNQOKOYICSPEDCKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHPASPSK 245
OY 181 RHAKAHEGYVCOGCSFVAKTWTELLKHVRETHKEELICEVCKRTFKRDYLLKOHKHTA 240
:|||||
Db 246 RHAKAHEGYVCOGCSFVAKTWTELLKHVRETHKEELICEVCKRTFKRDYLLKOHKHTA 305
OY 241 PERDYCRCPREGCGRTYTVFNLOSHILSFHESRPVCEHAGCGTFAMKOSLTRHAYV 300
:|||||
Db 306 PERDYCRCPREGCGRTYTVFNLOSHILSFHESRPVCEHAGCGTFAMKOSLTRHAYV 365
OY 301 HDPDKRKMRLKYKKSREKRSLSHLSGYIPPKRKGQGLSLCNGESPNCVEDKMLSTVA 360
:|||||
Db 366 HDPDKRKMRLKYKKSREKRSLSHLSGYIPPKRKGQGLSLCNGESPNCVEDKMLSTVA 425
OY 361 VLTG 365
|||||
Db 426 VLTG 430

RESULT 6
US-09-724-676-93267

; Sequence 93267, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93267
; LENGTH: 430
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-93267

Query Match 99.6%; Score 2000; DB 5; Length 430;

Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPYVAESVSLTTADAFIAGESSAPPPRPPALPRRFICFPDCCSANYSKAMKLDH 60
:|||||
Db 66 LDPYVAESVSLTTADAFIAGESSAPPPRPPALPRRFICFPDCCSANYSKAMKLDH 125
OY 61 LCKHTEGEPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
:|||||
Db 126 LCKHTEGEPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 185
OY 121 FERKHNQOKOYICSPEDCKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHPASPSK 180
:|||||
Db 186 FERKHNQOKOYICSPEDCKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHPASPSK 245
OY 181 RHAKAHEGYVCOGCSFVAKTWTELLKHVRETHKEELICEVCKRTFKRDYLLKOHKHTA 240
:|||||
Db 246 RHAKAHEGYVCOGCSFVAKTWTELLKHVRETHKEELICEVCKRTFKRDYLLKOHKHTA 305
OY 241 PERDYCRCPREGCGRTYTVFNLOSHILSFHESRPVCEHAGCGTFAMKOSLTRHAYV 300
:|||||
Db 306 PERDYCRCPREGCGRTYTVFNLOSHILSFHESRPVCEHAGCGTFAMKOSLTRHAYV 365
OY 301 HDPDKRKMRLKYKKSREKRSLSHLSGYIPPKRKGQGLSLCNGESPNCVEDKMLSTVA 360
:|||||
Db 366 HDPDKRKMRLKYKKSREKRSLSHLSGYIPPKRKGQGLSLCNGESPNCVEDKMLSTVA 425
OY 361 VLTG 365
|||||
Db 426 VLTG 430

RESULT 7
US-09-724-676-93278

; Sequence 93278, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93278
; LENGTH: 430
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-93278

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPYVAESVSLTTADAFIAGESSAPPPRPPALPRRFICFPDCCSANYSKAMKLDH 60
:|||||
Db 66 LDPYVAESVSLTTADAFIAGESSAPPPRPPALPRRFICFPDCCSANYSKAMKLDH 125
OY 61 LCKHTEGEPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
:|||||
Db 126 LCKHTEGEPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 185
OY 121 FERKHNQOKOYICSPEDCKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHPASPSK 180
:|||||
Db 186 FERKHNQOKOYICSPEDCKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHPASPSK 245
OY 181 RHAKAHEGYVCOGCSFVAKTWTELLKHVRETHKEELICEVCKRTFKRDYLLKOHKHTA 240
:|||||
Db 246 RHAKAHEGYVCOGCSFVAKTWTELLKHVRETHKEELICEVCKRTFKRDYLLKOHKHTA 305
OY 241 PERDYCRCPREGCGRTYTVFNLOSHILSFHESRPVCEHAGCGTFAMKOSLTRHAYV 300
:|||||
Db 306 PERDYCRCPREGCGRTYTVFNLOSHILSFHESRPVCEHAGCGTFAMKOSLTRHAYV 365


```

OY      241  PERDVCOPRECCGRTYTTVFMLOSHILSFHESRPVCEHAGCCGTFAMKOSTFRHAV 300
Db      306  PERDVCOCRCBSCGRTYTTVFMLOSHILSFHESRPVCEHAGCCGTFAMKOSTFRHAV 365
OY      301  HDPOKKKKKKLKKRSERKSLASHSGYTPPRKOGGSLCQNGESPVCVEDKMLSTVA 360
Db      366  HDPOKKKKKKLKKRSERKSLASHSGYTPPRKOGGSLCQNGESPVCVEDKMLSTVA 425
OY      361  VLTG 365
Db      426  VLTG 430

RESULT 11
US-09-724-676A-93223
; Sequence 93223, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93223
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93223

```

[illegible]

```

: SOFTWARE:PatentIn version 3.2
: SEQ ID NO.93234
: LENGTH: 430
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-724-676A-93234

Query Match          99.6%   Score 2000:  DB 5:   Length 430:
Best Local Similarity 99.5%:   Pred. No. 4e-176:
Matches 363:  Conservative 1:  Mismatches 1:  Indels 0:  Gaps 0:

```

Qy	1	MDPPVAVESVSLTADAFIAGESSAPTPRPALPREFICSPFDCSANYSKAKLDH	60
Db	66	LDPPVAVESVSLTADAFIAGESSAPTPRPLAPRFICSPFDCSANYSKAKLDH	125
Qy	61	LCKHGERPVCDEYEGCGKAFTRDYLSHLTLHTHGEKFPVCATGCDOKFPTKSNLKH	120
Db	126	LCKHGERPVCDEYEGCGKAFTRDYLSHLTLHTHGEKFPVCATGCDOKFPTKSNLKH	185
Qy	121	FERKHENOOKYICSEFEDCKTFKKHQOLKIQOCQHTNEPLFKTOEGCGKHAPSJKL	180
Db	186	FERKHENOOKYICSEFEDCKTFKKHQOLKIQOCQHTNEPLFKTOEGCGKHAPSJKL	245
Qy	181	RAAKHHEGVCCGCSFAKTTVELLKNVREHNKEELICEVCRKTFKKRDYLKQHNKTLA	240
Db	246	RAAKHHEGVCCGCSFAKTTVELLKNVREHNKEELICEVCRKTFKKRDYLKQHNKTLA	305
Qy	241	PERDVCRCPRECCGRTYTTVFMLQSHILSFHEBSRPVCEHAGCGGTFPMKOSLTHNAV	300
Db	306	PERDVCRCPRECCGRTYTTVFMLQSHILSFHEBSRPVCEHAGCGGTFPMKOSLTHNAV	365
Qy	301	HDPPDKMKMLKVKKSNEKRSLSHLSGYTPRKROGOGSLCQNGESPACVEDKMLSTVA	360
Db	366	HDPPDKMKMLKVKKSNEKRSLSHLSGYTPRKROGOGSLCQNGESPACVEDKMLSTVA	425
Qy	361	VLTG 365	
Db	426	VLTG 430	

```

RESULT 13
US-09-724-676A-93245
; Sequence 93245, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93245

Query Match          99.6%; Score 2000; DR 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0

QY      1 MDPPVVAVESVSLTIADAFIAAGESSAATPPRPALPRRFICSPDSCSANYSKAWKLDAN 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DBD     66 LDPPAVVAVESVSLTIADAFIAAGESSAATPPRPALPRRFICSPDSCSANYSKAWKLDAN 125

QY      61 LCKHHGERFVCDYGGCGCAFTRDVHLSNHLTHHGERFVCAANCGDCKFMTKSLKKH 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      126 LCKHHGERFVCDYGGCGCAFTRDVHLSNHLTHHGERFVCAANCGDCKFMTKSLKKH 185

QY      121 FERKHNQOKOYICGFEDCKKTFKKHQDLIHQCOHTNPLEFKCTOEGCGKHPASPSKLK 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      186 FERKHNQOKOYICGFEDCKKTFKKHQDLIHQCOHTNPLEFKCTOEGCGKHPASPSKLK 245

```

```

Qy 181 RHAKAHEGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNKOHMKTNA 240
      |||
Db 246 RHAKAHEGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNKOHMKTNA 305
Qy 241 PERDVCRCPREGCGRTYTVFNLOSHILSFHEESRPVCEHAGCGTFFAMKOSLTRHAYV 300
      |||
Db 306 PERDVCRCPREGCGRTYTVFNLOSHILSFHEESRPVCEHAGCGTFFAMKOSLTRHAYV 365
Qy 301 HDPDKKKMKLVKKSREKRSLSHLSGYIPPKRKOGGSLCQNGESPNCVEDKMLSTVA 360
      |||
Db 366 HDPDKKKMKLVKKSREKRSLSHLSGYIPPKRKOGGSLCQNGESPNCVEDKMLSTVA 425
Qy 361 VLTIG 365
      |||
Db 426 VLTIG 430

```

RESULT 14

```

US-09-724-676A-93256
; Sequence 93256, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OR INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93256
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93256

```

Query Match 99.6%; Score 2000; DB 5; Length 430;

Best Local Similarity 99.5%; Pred. No. 4e-176; Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 MDPRAVAESVSLTTADAFIAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 60
      :|||
Db 66 LDPRAVAESVSLTTADAFIAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 125
Qy 61 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLTHTGEPFVCAATGCDOKFNTKSNLKKH 120
      |||
Db 126 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLTHTGEPFVCAATGCDOKFNTKSNLKKH 185
Qy 121 FERKHNQOKOYICSFEDCKKTFKHQOLKHQCHTNEPLFKCTQEGCGKHFPASPSKLK 180
      |||
Db 186 FERKHNQOKOYICSFEDCKKTFKHQOLKHQCHTNEPLFKCTQEGCGKHFPASPSKLK 245
Qy 181 RHAKAHEGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNKOHMKTNA 240
      |||
Db 246 RHAKAHEGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNKOHMKTNA 305
Qy 241 PERDVCRCPREGCGRTYTVFNLOSHILSFHEESRPVCEHAGCGTFFAMKOSLTRHAYV 300
      |||
Db 306 PERDVCRCPREGCGRTYTVFNLOSHILSFHEESRPVCEHAGCGTFFAMKOSLTRHAYV 365
Qy 301 HDPDKKKMKLVKKSREKRSLSHLSGYIPPKRKOGGSLCQNGESPNCVEDKMLSTVA 360
      |||
Db 366 HDPDKKKMKLVKKSREKRSLSHLSGYIPPKRKOGGSLCQNGESPNCVEDKMLSTVA 425
Qy 361 VLTIG 365
      |||
Db 426 VLTIG 430

```

RESULT 15

```

US-09-724-676A-93267
; Sequence 93267, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OR INVENTION: Variants of alternative splicing

```

```

; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93267
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93267

```

Query Match 99.6%; Score 2000; DB 5; Length 430;

Best Local Similarity 99.5%; Pred. No. 4e-176; Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 MDPRAVAESVSLTTADAFIAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 60
      :|||
Db 66 LDPRAVAESVSLTTADAFIAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDAN 125
Qy 61 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLTHTGEPFVCAATGCDOKFNTKSNLKKH 120
      |||
Db 126 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLTHTGEPFVCAATGCDOKFNTKSNLKKH 185
Qy 121 FERKHNQOKOYICSFEDCKKTFKHQOLKHQCHTNEPLFKCTQEGCGKHFPASPSKLK 180
      |||
Db 186 FERKHNQOKOYICSFEDCKKTFKHQOLKHQCHTNEPLFKCTQEGCGKHFPASPSKLK 245
Qy 181 RHAKAHEGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNKOHMKTNA 240
      |||
Db 246 RHAKAHEGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNKOHMKTNA 305
Qy 241 PERDVCRCPREGCGRTYTVFNLOSHILSFHEESRPVCEHAGCGTFFAMKOSLTRHAYV 300
      |||
Db 306 PERDVCRCPREGCGRTYTVFNLOSHILSFHEESRPVCEHAGCGTFFAMKOSLTRHAYV 365
Qy 301 HDPDKKKMKLVKKSREKRSLSHLSGYIPPKRKOGGSLCQNGESPNCVEDKMLSTVA 360
      |||
Db 366 HDPDKKKMKLVKKSREKRSLSHLSGYIPPKRKOGGSLCQNGESPNCVEDKMLSTVA 425
Qy 361 VLTIG 365
      |||
Db 426 VLTIG 430

```

Search completed: February 10, 2003, 17:50:24
Job time : 42 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 12:58:51 ; Search time 70 Seconds
(without alignments)
694.807 Million cell updates/sec

Title: US-09-831-426c-2
Perfect score: 2008
Sequence: 1 MDPPAVVAEVSYSLTIDAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_101002:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	100.0	365	21	AAV93317
2	2000	99.6	409	22	AAV93317
3	1953	96.3	423	17	AAV93317
4	821	40.9	184	23	ABP41846
5	563.5	28.1	220	22	AAV93317
6	453	22.6	95	22	ABG09347
7	425	21.2	927	22	AAV93317
8	421	21.0	869	22	AAV93317
9	417.5	20.8	803	22	AAV93317
10	414.5	20.6	904	22	AAV93317

11	410.5	20.4	622	22	AAV93317	Human protein SEQ
12	410.5	20.4	631	22	AAV93317	Human protein SEQ
13	410.5	20.4	632	22	ABG18386	Novel human diago
14	410.5	20.4	652	22	AAV93317	Human protein SEQ
15	406.5	20.2	403	22	ABG64531	Drosophila melanog
16	406.5	20.2	700	22	AAV40001	Human polypeptide
17	405	20.2	1241	22	ABG02797	Novel human diago
18	404	20.1	779	22	ABG00880	Novel human diago
19	404	20.1	779	22	AAV93317	Human protein SEQ
20	404	20.1	779	22	AAV93317	Human protein SEQ
21	403	20.1	934	22	ABG69958	Drosophila melanog
22	398.5	19.8	773	22	AAV41294	Human polypeptide
23	396.5	19.7	646	22	ABP51377	Human MDPF SEQ ID
24	396	19.7	577	22	AAV93317	Human polypeptide
25	395.5	19.7	412	21	AAV93316	A transcription fa
26	395.5	19.7	1050	22	ABG09685	Novel human diago
27	395	19.7	803	22	AAV93278	Human protein sequ
28	394.5	19.6	727	22	ABG16954	Novel human diago
29	394.5	19.6	809	22	AAV93689	Human polypeptide
30	393.5	19.6	751	22	AAV93508	Human polypeptide
31	393	19.6	489	22	ABG50238	Human transcription
32	393	19.6	702	23	ABG04333	Human zinc finger
33	393	19.6	799	21	AAV21003	Human nucleic acid
34	392.5	19.5	547	22	AAV78366	Human protein SEQ
35	392.5	19.5	570	22	AAV79350	Human protein sequ
36	391.5	19.5	632	22	ABG19011	Novel human diago
37	388.5	19.3	817	22	AAV40475	Human polypeptide
38	387.5	19.3	1252	22	AAV79739	Human protein SEQ
39	386.5	19.2	759	23	AAV78844	Mouse SCAN/KRAB pr
40	386	19.2	406	22	AAU16023	Human novel secret
41	385.5	19.2	577	23	AAE16023	Human transcriptio
42	385.5	19.2	1520	22	ABG27130	Novel human diago
43	384.5	19.1	357	22	AAV93635	Human protein sequ
44	384.5	19.1	386	22	AAV93303	Human polypeptide
45	384.5	19.1	613	22	AAV93862	Human protein sequ

ALIGNMENTS

RESULT 1
ID AAV93317 standard; Protein: 365 AA.
AC AAV93317;
DT 04-SEP-2000 (first entry)
DE A human transcription factor designated htflf1A.
KW Human; transcription factor; htflf1A; DNA-binding protein;
KW Transcription; ribosomal RNA 5S gene; transcriptional control;
KW cancer.
XX
XX Homo sapiens.
OS
XX
XX WO200028024-A1.
PD 18-MAY-2000.
PF 09-NOV-1999; 99WO-FR02738.
PR 10-NOV-1998; 98FR-001446.
XX
XX (HMRI) HOECHST MARION ROUSSEL.
XX Bordon-Pallier F, Roher C;
XX
XX WPI: 2000-387419/33.
XX N-PSDB; AAA15405.
XX
XX New nucleic acid encoding human transcription factor flf1A, useful for
PT treatment and diagnosis of cancer and inherited disease -

XX Claim 10; Page 40-41; 49pp; French.
 CC The present sequence represents a human transcription factor designated
 CC htf11a. The polypeptide is probably a DNA-binding protein probably
 CC involved in initiating transcription of the gene for ribosomal RNA 5S
 CC and maintaining the stability of transcription of other control genes.
 CC The htf11a polynucleotides and polypeptides are used to make
 CC therapeutic or diagnostic compositions for diseases associated with
 CC disorders of transcriptional control, particularly cancer or other
 CC inherited diseases. The htf11a polynucleotide can also be used to
 CC detect anomalies in gene transcription, particularly for diagnosis of
 CC inherited disease, also for studying diseases involving htf11a.
 CC
 XX Sequence 365 AA:
 SQ
 Query Match 100.0%; Score 2008; DB 21; Length 365;
 Best Local Similarity 100.0%; Pred. No. 2,56-159;
 Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDPRAVAVSVSLTADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDH 60
 DB 1 MDPRAVAVSVSLTADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDH 60
 QY 61 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDQKFNKSNLKKH 120
 DB 61 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDQKFNKSNLKKH 120
 QY 121 FERKHNQOKYITCSFEDCKTFKKHQDLKHQCCHTNPELFCQDEGCGKHAPSFKLK 180
 DB 121 FERKHNQOKYITCSFEDCKTFKKHQDLKHQCCHTNPELFCQDEGCGKHAPSFKLK 180
 QY 181 RHAKAHEGYVCGKGFVAKTWTTELKHYRETHKEEILCEVCRKTKFRDYLKQNHKTHA 240
 DB 181 RHAKAHEGYVCGKGFVAKTWTTELKHYRETHKEEILCEVCRKTKFRDYLKQNHKTHA 240
 QY 241 PERDYCRCPREGCGRTYTTVFNLQSHILSFHESSRPFCVCAATGCDQKFNKSNLKKH 300
 DB 241 PERDYCRCPREGCGRTYTTVFNLQSHILSFHESSRPFCVCAATGCDQKFNKSNLKKH 300
 QY 301 HDPDKKKMLKVKSKREKSLASHLSGYIPPKKQGGSLCQNGESPNCVEDKMLSTVA 360
 DB 301 HDPDKKKMLKVKSKREKSLASHLSGYIPPKKQGGSLCQNGESPNCVEDKMLSTVA 360
 QY 361 VLTIG 365
 DB 361 VLTIG 365
 RESULT 2
 AAG75181
 ID AAG75181 standard; Protein: 409 AA.
 XX
 AC AAG75181;
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:5945.
 XX
 KM Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 13.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000MO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI: 2001-235357/24.
 XX
 DR N-PSDB: AAH34586.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 XX useful for preventing, diagnosing and/or treating colorectal cancers -
 PS Claim 11; Page 7452-7454; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 409 AA:
 Query Match 99.6%; Score 2000; DB 22; Length 409;
 Best Local Similarity 99.5%; Pred. No. 1,4e-158;
 Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MDPRAVAVSVSLTADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDH 60
 DB 45 LDPRAVAVSVSLTADAFIAAGESSAPTPRPALPRRFICSPDCSANYSKAMKLDH 104
 QY 61 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDQKFNKSNLKKH 120
 DB 105 LCKHTGERPFCVDEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDQKFNKSNLKKH 164
 QY 121 FERKHNQOKYITCSFEDCKTFKKHQDLKHQCCHTNPELFCQDEGCGKHAPSFKLK 180
 DB 165 FERKHNQOKYITCSFEDCKTFKKHQDLKHQCCHTNPELFCQDEGCGKHAPSFKLK 224
 QY 181 RHAKAHEGYVCGKGFVAKTWTTELKHYRETHKEEILCEVCRKTKFRDYLKQNHKTHA 240
 DB 225 RHAKAHEGYVCGKGFVAKTWTTELKHYRETHKEEILCEVCRKTKFRDYLKQNHKTHA 284
 QY 241 PERDYCRCPREGCGRTYTTVFNLQSHILSFHESSRPFCVCAATGCDQKFNKSNLKKH 300
 DB 285 PERDYCRCPREGCGRTYTTVFNLQSHILSFHESSRPFCVCAATGCDQKFNKSNLKKH 344
 QY 301 HDPDKKKMLKVKSKREKSLASHLSGYIPPKKQGGSLCQNGESPNCVEDKMLSTVA 360
 DB 345 HDPDKKKMLKVKSKREKSLASHLSGYIPPKKQGGSLCQNGESPNCVEDKMLSTVA 404
 QY 361 VLTIG 365
 DB 405 VLTIG 409

RESULT 3
 AAR91305
 ID AAR91305 standard; Protein: 423 AA.
 XX
 AC AAR91305;
 XX

DT	07-JUL-1996	(first entry)
XX		
DE	Transcription factor-IIIA.	
XX		
KW	Human; transcription factor-IIIA; hTFIIIA; DNA binding protein;	
KV	ribosome; zinc finger; diagnostic; probe; transcription control;	
KW	antitumour; cancer; therapy.	
XX		
OS	Homo sapiens.	
XX		
PN	EP704526-A1.	
XX		
PD	03-APR-1996.	
XX		
PF	05-SEP-1995; 95EP-0113908.	
XX		
PR	05-SEP-1994; 94JP-0211022.	
XX		
PA	(SAKA) OTSUKA PHARM CO LTD.	
XX		
PI	Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;	
DR	WPJ: 1996-173033/18.	
DR	N-PSTD: AAT14037, AAT14038.	
PT	Human Transcription Factor III A gene - useful in regulation of	
PT	transcription and for diagnosis and treatment of e.g. cancer related	
PT	diseases	
XX		
PS	Claim 1; Page 8-10; 17pp; English.	
XX		
CC	The sequence represents human transcription factor-IIIA (hTFIIIA),	
CC	a DNA binding protein which is necessary for the initiation of 5S	
CC	rRNA gene transcription, binding to an internal control region of	
CC	the 5' gene. The protein contains 9 zinc finger domains, which are	
CC	homologous to the C2H2 finger domains of Xenopus TFIIIA, except for	
CC	the 6th finger domain, which has only 3 amino acid residues between	
CC	2 cysteine residues, instead of 5 amino acid residues for Xenopus	
CC	TFIIIA. The protein (optionally in recombinant form) and encoding	
CC	gene may be used in diagnosis, identification or therapy of	
CC	hereditary diseases such as cancer, or other diseases resulting from	
CC	abnormal transcriptional control, and to analyse the mechanisms	
CC	involved in their activity.	
XX		
SQ	Sequence 423 AA:	
	Query Match 96.3%; Score 1933; DB 17; Length 423:	
	Best Local Similarity 96.2%; Pred. No. 5; Se-153;	
	Matches 351; Conservative 3; Mismatches 11; Indels 0; Gaps	
OY	1 MDPPAVVAESVSLTIADFFIAGESSAPTTPRRPALPFRFICSPDCSANYSKAWLDAH 60	
	: : : : : : : : :	
DB	59 LDPRAVVAESVSLTIADFFIAGESSATTPRRPALPFRFICSPDCSANYSKAWLDAH 118	
OY	61 LCKHTGERPEVCDFEGCGKAFTRDYHLNSHLLTHGGEKFPVCAATGCDOKFNTRSNLKKH 120	
	: : : : : : : : :	
DB	119 LCKHTGERPEVCDFEGCGKAFTRDYHLNSHLLTHGGEKFPVCAANGCDOKFNTRSNLKKH 178	
OY	121 PERKHENQOKVICSEEDCKTKFKKNQOLKINOCQNTNPERLKRCTDGGCGKHFASSSKLR 180	
	: : : : : : : : :	
DB	179 FERKHENQOKVICSEEDCKTKFKKNQOMIKIQCOQTNPRLFKCTDGGCGKHFASSSKLR 238	
OY	181 RRAKNAHEGVCCGSGSFVAKTWTLLKNAHRETNKEEILDEVCRKTRKRDYLKONHKHTA 240	
	: : : : : : : : :	
DB	239 RRAKNAHEGVCCGSGSFVAKTWTLLKNAHRETNKEEILDEVCRKTRKRDYLKONHKHTA 298	
OY	241 PERDVCCRCREGCGRTYTTFVNLSHIISFHESRPFCVEHAAGCGKTAMKOSLTRHAVV 300	
	: : : : : : : : :	
DB	299 PERDVCCRCREGCGRTYTTFVNLSHIISFHESRPFCVEHAAGCGKTAMKOSLTRHAVV 358	
OY	301 HDPDKKKMLKLYKRSRKSLASHLSGYTPPKRGQGGLSLCNGESPNCVEDKMLSTYA 360	
	: : : : : : : : :	
DB	359 HDPDKKKMLKLYKRSRKREKFGLSSOWITYPKRGQGGLSLCNGESPNCVEDKMLSTYA 418	

OY	361	VLTIG	365
Db	419	VLTIG	423
RESULT 4			
ID	ABP41846	ABP41846 standard; Protein; 184 AA.	
AC	ABP41846;		
XX			
DT	22-AUG-2002	(first entry)	
DE	Human ovarian antigen HUSYA18,	SEQ ID NO:2978.	
KW	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;		
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;		
KW	infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;		
KM	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;		
KW	Inflammatory condition; immune disorder; blood disorder;		
KW	cardiovascular disorder; respiratory disorder; neurological disorder;		
KW	gastrointestinal disorder; urinary system disorder; drug screening;		
KW	gene therapy; chromosome mapping; forensic analysis;		
KM	antibody preparation; cytostatic; immunomodulatory; neuroprotective;		
KM	antiinflammatory; gynaecological; reproductive; chromosome 13q12.3-13.1.		
OS	Homo sapiens.		
XX			
PN	WO200200677-A1.		
PD	03-JAN-2002.		
XX			
PE	07-JUN-2001; 2001WO-US18569.		
XX			
PR	07-JUN-2000; 2000US-209467P.		
XX			
PA	(HOMA-) HUMAN GENOME SCI INC.		
PI	Birse CE, Rosen CA:		
XX			
DR	WPI: 2002-147878/19.		
XX	N-PSDB; ABQ54923.		
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,		
PT	useful in the prevention, treatment and diagnosis of cancer (e.g.		
PT	ovarian cancer), immune disorders, cardiovascular disorders and		
PT	neurological diseases -		
XX			
PS	Claim 11; SEQ ID No 2978; 2972pp; English.		
CC			
XX	The invention relates to 2175 novel human ovarian antigens (ABP41054-		
CC	ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also		
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical		
CC	to the sequences of the invention. The invention additionally relates to		
CC	recombinant vectors and host cells comprising human ovarian antigen		
CC	polynucleotides, antibodies against human ovarian antigens, and the use		
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,		
CC	treating, prognosing or preventing various ovary and/or breast-related		
CC	disorders. Such conditions include ovarian cancer and breast cancer, and		
CC	metastatic tumours of ovarian or breast origin, reproductive system		
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,		
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine		
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic		
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and		
CC	vaginitis), immune disorders (e.g., congenital and acquired		
CC	immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),		
CC	blood-related disorders (e.g., anaemia), cardiovascular disorders,		
CC	respiratory disorders, neurological disorders, gastrointestinal disorders		
CC	and urinary system disorders. Ovarian antigen polypeptides which		
CC	polynucleotides may also be used in screening for compounds and		
CC	modulate ovarian antigen expression or activity. The polynucleotides may		
CC	further be used for gene therapy, chromosome mapping, in the		

PT	biodiversity -
XX	
PS	Claim 20; SEQ ID No 39706; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pcr_sequences.
XX	
SQ	Sequence 95 AA;
	Query Match 22.6%; Score 453; DB 22; Length 95;
	Best Local Similarity 91.6%; Pred. No. 2.4e-30;
	Matches 87; Conservative 1; Mismatches 7; Indels 0; Gaps 0
QY	271 HEENRPFVCEHAGCGCKTYRAMKSLTRHAHVADPPDKMKMLKYKKSEKRSLASHLSGYIP 330 : PFVCEHAGCGCKTFRANKPSLSDAVVHDPGKKMKMLKYKKSEKRSLASHLSGYIP 60
Dd	1 HEENRPFVCEHAGCGCKTFRANKPSLSDAVVHDPGKKMKMLKYKKSEKRSLASHLSGYIP 60
QY	331 PKRKGGGLSLCQNGESPNCVEDKMLSTVAIVLTG 365 PKRKGGGLSLCQNGESPNCVEDKMLSTVAIVLTG 95
Dd	61 PKRKGGGLSLCQNGESPNCVEDKMLSAVPVLTG 95
	RESULT 7
ID	AAAM80283 standard; Protein; 927 AA.
XX	
AC	AAAM80283;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human protein SEQ ID NO 3929.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorder; arthritis; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	WO200157190-A2.
XX	
PD	09-AUG-2001.
XX	
PF	05-FEB-2001; 2001WO-US04098.
XX	
PR	03-FEB-2000; 2000US-0496914.
PR	27-APR-2000; 2000US-0560875.
PR	20-JUN-2000; 2000US-0598075.
PR	19-JUL-2000; 2000US-0620325.
PR	01-SEP-2000; 2000US-0654936.
PR	15-SEP-2000; 2000US-0663561.
PR	20-OCT-2000; 2000US-0693325.
PR	30-NOV-2000; 2000US-0728422.
XX	

(HXSE-) HXSEO INC.

XX Tang YT, Liu C, Dirmacac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WP1: 2001-476283/51.
DR N-PSDB; AAK53416.

XX

PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -

XX
XS
FS
XX Claim 20; Page 470-471; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation on which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK5281), 2111 (AAK5282) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

[illegible]

```

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0634936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
XX N-PSDB: AAK52432.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX Claim 20: Page 4372-4374; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX Sequence 869 AA:
XX
XX Query Match 21.0%; Score 421; DB 22; Length 869;
XX Best Local Similarity 32.0%; Pred. No. 1.8e-26;
XX Matches 116; Conservative 41; Mismatches 116; Indels 89; Gaps 19;
OY 3 PRAVVAESVSLTIADAFIAAGESSAPTPRPA-LPRRTICSPDCAANTYKAMKIDAH 61
DB 125 PVPVVAEPNERSLRELVGG-----RPAAGAEKPTICN--ECGKSFSSOMSKILRHQ 169
OY 62 CHTGGERPVCYEGCGKAFIRDYHLNRHLITHTGKRPVC----- 102
DB 170 RHTGERPHTC--SECGKFTOSSHLVQHQRHTTGKPKPCDCGCGFSWSSNLVQHQT 227
OY 103 -----AATGDDOKFNTSKLNKHEERKHNQOQOYLISFEDCKKTFKKHOOLKIHCO 155
DB 228 HTGGEPRYKCTCEELAFSTQSTNLIKH-QRSHTG-EKPYKCG--ECRAFAFRSSDLQHQT 283
OY 156 HINPELTKTOGCGCKHAFSPSKLRHAKAHAG--YVQCK-GCSPV-----AKTWT 203
DB 284 HNGEPRYKCP--CGKRFQONHNLKHKRIHAGEKPRCTECGKSFIOSESLTQHQRHTT 341

```

```

OY 204 -----ELTKHVRETHKE--ILCEVCRTFKRDYLKOHKTHAPER 243
DB 342 GEPHYCLECGKSGFHSSTLIKHR--THLRDEPFRKPCVCKFTTSLATLRLRHQRHTTGER 400
OY 244 DVCRCPRGCGRTYTTVFNIOSHILSFHEBSRPFVCEHAGCGTFAWKOSLTRHAVVDP 303
DB 401 PY-KCPE--CGKSFVSNSLNH-QRIHGERPYIC--ADCGKSFIMSSLTIRHQRIHTG 454
OY 304 DK 305
DB 455 EK 456
RESULT 9
AAM79014
ID AAM79014 standard. Protein; 803 AA.
XX AAM79014;
XX 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 1676.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0634936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
XX N-PSDB: AAK52147.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX Claim 20: Page 4019-4021; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing

```


PR 30-NOV-2000: 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y:
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW:
PI Xue AJ, Yang Y, Weighman T, Goodrich R;
XX WPI: 2001-476283/51.
DR N-PSDB: AAK52304.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20: Page 4219-4221: 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK5281), 2111 (AAK5282) and 3666
CC (AAW80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 652 AA:
XX
Query Match 20.4%; Score 410.5; DB 22; Length 652;
Best Local Similarity 37.7%; Pred. No. 9.5e-26;
Matches 100; Conservative 31; Mismatches 111; Indels 23; Gaps 14;
OY 46 DCSANYSKAWKLDALHCKHTGERPVCDEGCGKAFIRDYHLSRHLTHTGKRPVCAAT 105
DB 375 ECGKAFWTSSLNKHKRHTGKPYTC--EECGKAFYRSHLAKKRITGKPTCE- 431
OY 106 GCDQFNFKSNLKKFERKHNENQOYICSPEDCKTKFKHOOOLKHOCQHTNEPLFKCT 165
DB 432 -CGKAFNOSSTLILH-KRIHSG-QKPYKC--EECGKAFRSTLTNEHKKIHTGKPYKC- 485
OY 166 QEGCKNHFPSPKSLKRNAAHNEG---YVOCQKCSFVAKTWTLLKHVRETKEIL--CE 220
DB 486 -EECGKAFITMSALNENKNIHTGKPYKC-KECGKAFNOSGLIITN-NSIHSEQILKYCE 542
OY 221 VCRKTFKRDYLNQHMKTTHAPRDYCCRCRREGCGRTYTTVFNLQSHILSFHESRPVCE 280
DB 543 ECGKAFTRSTALNENKNIHTGKPYKC-KC--KECGKAYNLSTLTGN-KRIHTGKRPVCE 598
OY 281 HAGCGKTFPMKQSLTRHNAVHDPDK 305
DB 599 E--CGKAFNMSSSLTKHKIHTGKEK 621
XX
RESULT 15
ABW64531
ID ABW64531 standard; Protein: 403 AA.
XX
XX ABB64531;
AC
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 20385.
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 20385.
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX

PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001: 2001WO-US09231.
XX
XX 23-MAR-2000: 2000US-191637P.
PR 11-JUL-2000: 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR N-PSDB: ABL08634.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure: SEQ ID NO 20385; 21pp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16175-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABW57737-ABW2072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 403 AA:
XX
Query Match 20.2%; Score 406.5; DB 22; Length 403;
Best Local Similarity 30.3%; Pred. No. 1.1e-25;
Matches 105; Conservative 55; Mismatches 144; Indels 43; Gaps 14;
OY 39 RTGCFPPDCSANYSKAWKLDALHCKHTGERPVCDEGCGKAFIRDYHLSRHL-LTH--- 94
DB 34 KYACSPKPCGATFKRLDQDRHEYNHTGIKKACSYEGCDKYSIVTLKRLRLSTHERP 93
OY 95 --TGKFPVCAATGCDQFNFKSNLKKFERKHNENQOYICSPEDCKTKFKHOOOLKH 152
DB 94 ESAAKTYKCALEBSCSKAFISVSNMTRHRETRES-PKVTYPCS--QCSAKFSQKLKRN 150
OY 153 QC-QHTNEPLFKCTQEGCGKHFASPSKLRNAAHNEGYYCQKCSFVAKTWTLLKHVRE 211
DB 151 EIREHTLEVPYCSK--CSRGFYQOMOCOSHESPCKLYEC-PGCPLOPDKMTLYTKHCD 207
OY 212 T-----HKELLCVCKRTFKRKDYLNQHMKT---HAPERVVC---RCPRREGCGRTT 258
DB 208 SLHGKRNRRK---CDRCDSAFDPSKRLNHLKVKHKAQVDECATSFTCNEGGGKSYIS 263
OY 259 TVFNLSHILSFHESRPVCEHAGCGKTFAMKQSLTRHNAV-VHPRDKKMKLYKKKS 317
DB 264 YLRNLKQHMILTAN-SGRRECCALDCCGRCFSSAQMLARILRLDHDKATKKLAKKKDK 322
OY 318 KRSLSHLSGYIPPRKQGGISLCQNGESPNCVDEKMLSTVAVLTL 364
DB 323 SKTGGGKTKSTSRKRRRDAGRS-----KHSRLSKLACQL 358
XX

Search completed: February 10, 2003, 17:45:15
Job time : 77 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:44:27 ; Search time 24 Seconds
(without alignments)
447.474 Million cell updates/sec

Title: US-09-831-426C-2
Perfect score: 2008
Sequence: 1 MDPFAVAESVSLITADAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/CTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1933	96.3	423	1	US-08-523-376-3 Sequence 3, Appli
2	381	19.0	711	2	US-08-820-170A-10 Sequence 10, Appl
3	381	19.0	711	3	US-09-055-699-10 Sequence 10, Appl
4	381	19.0	711	4	US-09-273-565-10 Sequence 10, Appl
5	381	19.0	711	4	US-09-565-538-10 Sequence 10, Appl
6	381	19.0	711	4	US-09-661-468-10 Sequence 10, Appl
7	353.5	17.6	543	4	US-09-362-123A-4 Sequence 4, Appli
8	336	16.7	803	4	US-09-063-035-2 Sequence 2, Appli
9	318.5	15.9	727	2	US-08-475-844-9 Sequence 9, Appli
10	318.5	15.9	727	2	US-08-475-844-9 Sequence 9, Appli
11	313	15.6	640	4	US-09-262-773-4 Sequence 2, Appli
12	313	15.6	640	4	US-09-262-773-4 Sequence 2, Appli
13	312.5	15.6	462	3	US-08-486-099-117 Sequence 117, App
14	312.5	15.6	462	3	US-08-360-107A-127 Sequence 127, App
15	312.5	15.6	462	3	US-08-484-223B-117 Sequence 117, App
16	312.5	15.6	462	3	US-08-919-597-117 Sequence 117, App
17	312.5	15.6	462	3	US-08-475-668A-117 Sequence 117, App
18	312.5	15.6	462	3	US-08-485-551A-117 Sequence 117, App
19	312.5	15.6	462	3	US-08-471-913A-117 Sequence 117, App
20	312.5	15.6	462	4	US-08-485-264A-117 Sequence 117, App
21	312.5	15.6	462	4	US-08-474-349A-117 Sequence 117, App
22	312	15.5	728	2	US-08-475-844-5 Sequence 5, Appli
23	312	15.5	728	2	US-08-475-844-5 Sequence 5, Appli
24	301.5	15.0	488	2	US-08-933-750C-17 Sequence 17, Appl
25	301.5	15.0	488	2	US-09-234-613-17 Sequence 17, Appl
26	288.5	14.4	181	4	US-08-676-318A-44 Sequence 44, Appl
27	278.5	13.9	181	4	US-08-863-813A-44 Sequence 44, Appl

28	259.5	12.9	667	2	US-08-718-661-2 Sequence 2, Appli
29	251.5	12.5	441	4	US-09-172-045-2 Sequence 2, Appli
30	248	12.4	345	1	US-08-102-942A-2 Sequence 2, Appli
31	248	12.4	345	4	US-09-037-179B-2 Sequence 2, Appli
32	246	12.3	429	1	US-08-234-783-4 Sequence 4, Appli
33	246	12.3	429	1	US-08-456-907-4 Sequence 4, Appli
34	246	12.3	429	5	PCR-US95-05523-4 Sequence 4, Appli
35	243.5	12.1	449	1	US-08-102-942A-4 Sequence 4, Appli
36	243.5	12.1	449	4	US-09-037-179B-4 Sequence 4, Appli
37	242.5	12.1	449	3	US-09-234-332-11 Sequence 11, Appl
38	242.5	12.1	449	1	US-08-102-942A-6 Sequence 6, Appli
39	242.5	12.1	449	4	US-09-037-179B-6 Sequence 6, Appli
40	238	11.9	338	2	US-08-933-750C-4 Sequence 4, Appli
41	238	11.9	338	4	US-09-234-613-14 Sequence 4, Appli
42	224.5	11.2	671	3	US-09-121-321-16 Sequence 16, Appl
43	224.5	11.2	671	4	US-08-933-803A-16 Sequence 16, Appl
44	222	11.1	414	5	PCR-US92-06840-2 Sequence 2, Appli
45	221	11.0	706	1	US-08-074-967-2 Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-523-376-3
; Sequence 3, Application US/08523376
; Patent No. 5808030
;
GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMURA
; APPLICANT: Satoshi, TAKEIDA
; APPLICANT: Yoshikazu, SHIMADA
; APPLICANT: Kouichi, OZAKI
; APPLICANT: Sadahito, SIN
; TITLE OF INVENTION: htf11A GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,376
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ. ID NO.: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-523-376-3

Query Match          96.3%   Score 1933;   DB 1;   Length 423;
Best Local Similarity 96.2%   Pred. No. 8.3e-189;
Matches 351;   Conservative 3;   Mismatches 11;   Indels 0;   Gaps 0;

OY      1  MDPFAVAESVSLITADAFITAGESSAPPPPPALPFRFICSPPOCSANYSAKMLDAH 60
Db      59  LDPPFAVAESVSLITADAFITAGESSAPPPPPALPFRFICSPPOCSANYSAKMLDAH 118
OY      61  LCRHTGERPVCDEGCGKAFIRDYHLIRHILTHTGERKPPVCAATGCDKFNTRSKNLKH 120
|||||
```


Query Match 19.0% Score 381; DB 4; Length 711;
Best Local Similarity 32.9% Pred. No. 2.6e-30;
Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;

OY 38 RRTICSPFDCSANSYKAMKIDHLCKHTGERPVCYEGCGKAFIDYLSHILITHTGE 97
DB 445 KSYVCI--EGGAFIQKAMHLYHQRHTGKERPYOC--HHGCGKSFISKQSDIHHRHTGE 500
OY 98 KPVCAATGCDGCKFNTKSNLKKHFERKHENQOQYICSPEDCKTKTKFKHQOLKHOCQHT 157
DB 501 KPEEC--SDCGKFTQKSHLNTIH--QKHTGEMHWCS--EGCGAFNQSLISMHORHT 554
OY 158 NEPLFCTGEGCGKHFASPSKLKRNKAHEG---YVCGKCSFVAKTWTELLKH-VRETH 213
DB 555 GCKPYKCS--CGKAFYSQSOFKHOIRHTGKERPYCTE--CGKAFGRSNFHKHQTHTIR 611
OY 214 KEIILCEVCKTKTKKRDYLYKOHKTHAFEDVCKPRE--GCGRTTYTFVNLQSHILSTR 271
DB 612 ERPFVCKGCKAFVQKSELTHQRTHMGK---PEGLDCKGKSPSKRPOLKVH-QRTH 665
OY 272 EESRPVCEHAGCGKTFAMKOSLTRHAYVHDPDK-KMKLVK 313
DB 666 TGERPYVCSF--CGKAFNNRBNFNKHQHTHTTRDKSYKCSYSK 706

RESULT 7

US-09-362-123A-4
; Sequence 4, Application US/09362123A
; Patent No. 6451558
; GENERAL INFORMATION:
; APPLICANT: Cooke, Michael Paul
; APPLICANT: Holness, Claire
; APPLICANT: Sirenko, Oksana
; TITLE OF INVENTION: No. 6451558el Genes in the Control of Hematopoiesis
; FILE REFERENCE: 4-30629A/SYS
; CURRENT APPLICATION NUMBER: US/09/362,123A
; CURRENT FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: 09/128,310
; PRIOR FILING DATE: 1998-08-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-362-123A-4

Query Match 17.6% Score 353.5; DB 4; Length 543;
Best Local Similarity 31.5% Pred. No. 1.2e-27;
Matches 100; Conservative 38; Mismatches 144; Indels 35; Gaps 15;

OY 23 AGSSAPTPRRALPRRIFCSFPDCSANSYKAMKIDHLCKHTGERPVCYEGCGKAFI 82
DB 242 AGKKS-----FEER--ACSKVFKVSSDLKHLKRTHTGKERPYCC--AQCGKAFS 285
OY 83 RYHLSRHILITHTGKERPYCAATGCDGCKFNTKSNLKKHFERKHENQOQYICSPEDCKRT 142
DB 286 QSHLHQHQRHNSGEPYACPV--CGKAFRHSLSLVRI-QRIH-TAKESFRCS--EGCGA 339
OY 143 FKKHQOLKHOCQHTNEPLFCTGEGCGKHFASPSKLKRNKAHEG---YVCGKCSFVA 199
DB 340 FSHGSMISOHRIKTHAGGRYACAO--CGRRFCNSHLIOHERHTHTGKERPYCAL--CGAAF 396
OY 200 KTMTELLKIVR-ETHKEIILCEVCKTKTKKRDYLYKOHKTHAFEDVCKPREGCGRTT 258
DB 397 SOGSSLFKIHQRHNTGKERPRACPCGGRFSSNSLTQHOLLHTGERPF-RC--VDCKAKFA 453
OY 259 TVFNLOSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAYVHDPDKKMKLVKKSREK 318
DB 454 KQAVLLSH-RRIHTGKERPYCTO--CGRAFRERPALFHHORHTHTGKTYRBSASLHPQA 510
OY 319 RSLASHLSGYIPPKRKQ 335

DB 511 RSVAGASSEGAPEKETE 527

RESULT 8

US-09-063-035-2
; Sequence 2, Application US/09063035
; Patent No. 6160091
; GENERAL INFORMATION:
; APPLICANT: PEUKERT, Karen; HAENEL, Frank; and EILERS,
; TITLE OF INVENTION: Myc-binding zinc finger proteins,
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.1
; SOFTWARE: Wordperfect version 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,035
; FILING DATE: 21-APR-1998
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 803 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-063-035-2

Query Match 16.7% Score 336; DB 4; Length 803;
Best Local Similarity 31.8% Pred. No. 1.2e-25;
Matches 95; Conservative 45; Mismatches 117; Indels 42; Gaps 16;

OY 52 SKAMKIDHLCKHTGERPVCYEGCGKAFIRDYLSHILITHTGKERPVCAATGCDOKF 111
DB 298 SKAYGSVIRKC-----EDCGKEFTHTGNFKRRIHTHTGKPPSCRE--CSKAF 343
OY 112 NTKSNLKKHFERKHENQOQYICSPEDCKTKTKKHOOLKHOCQHTNEPLFCTGEGCGK 171
DB 344 SDPACKAH-EXTH-SPLKPYGC--EGCGSYRLISLNLKHKRHSGEARYRC--EDCGK 397
OY 172 HFASPSKLKRNKAHEG---YVCGKCSFVAKTWTELLKHVRETH--KEIILCEVCKRTF 226
DB 398 LTTTSGNLKRHLVHSGEPYOCY--CGRSFSDPRTSKMHL-ETHDTQKHNKCPICDKKF 455
OY 227 KKKDYLYKOHKTHAFEDV-CRCPRGCGRTTYTFVNLQSHILSFHEESRPVCEHAGCG 285
DB 456 NOVGSLKAMKIDHLIDGRLKCR---ECGKOPTTSGNLKRO-LRIHSGEPYVCIH--CQ 508
OY 286 KTFAMKOSLTRHAYVHDPDKKMKLVKKS-REKRSLSHLSGYIPPKKQOGLSLQ 343
DB 509 RQFADGALORHVRIHTHTGKERPCQCVWCGKAFYTOASSLAHV-----RQHTGKERPYCE 561
RESULT 9
US-08-475-844-9
; Sequence 9, Application US/08475844
; Patent No. 5972643
; GENERAL INFORMATION:
; APPLICANT: Lobanenkoy, Victor V.
; APPLICANT: Klenova, Elena M.
; APPLICANT: Goodwin, Graham H.
; APPLICANT: Filippova, Galina N.
; APPLICANT: Collins, Steven J.


```

: APPLICANT: Hess, Mark A.
: TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
: FILE REFERENCE: Myriad 3
: CURRENT APPLICATION NUMBER: US/09/262,773
: CURRENT FILING DATE: 1999-03-04
: NUMBER OF SEQ ID NOS: 210
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 640
: TYPE: PRT
: ORGANISM: human
US-09-262-773-4

```

```

Query Match          15.6% Score 313; DB 4; Length 640;
Best Local Similarity 31.6% Pred. No. 1.9e-23;
Matches 95; Conservative 44; Mismatches 122; Indels 40; Gaps 17;

```

```

QY 12 SSUTLADAFIAGESSAPPPRALPRRFICSPDCSANSKAMKLDALHCKHTGERPFV 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 TNISQVNSFVNLRET---TPVHPLGRHNDGSY--CGKSTCSHLVRLHRTHTGEKPYK 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 CDYEGCGKAFIRDYHLNRHILHTTGKPRVCAATGCDQKFN--TKSNLKK-----HEERKH 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 C--MECGKSTYRSSHLARHOKVHKMNAFY-----KYLNRKMLEETSPVTOAERT- 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 ENQOKOYICSEFDECKTFKKHQOLKIHQOQTNELPKCTOEGCGNHFASPSKLKRNHAKA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 467 PSVEKPYRC--DDCGKHFMTSDLVNRHQRTHTEKPFCTI--CGKSFQSKSVLTTHQRI 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 HEG---YVCOKCGSFVAKTWTLLKHVRETH--KEELICEVCRKTKRKDYLRKQHKHTHA 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 HUGGKPYLIGE--CGEDFSEHRRYLAH-RKTHAAEELYLCEGCRCTHSAAFKHLRGHA 580
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 PERDVCRCPREGCGRTYTVFNLOSILSFHESRPPVCEHAGCGKTFAMKOSLTRHAYV 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 581 SVAP-CRC--NECGKSFSRDHLVRHQRTHTEKPFCTC--PTCGKSFSGYHLIRHQR 634
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 H 301
   :
Db 635 H 635

```

```

RESULT 12
US-09-262-773-2
: Sequence 2, Application US/09262773
: Patent No. 6225451
: GENERAL INFORMATION:
: APPLICANT: Ballinger, Dennis G.
: APPLICANT: Ding, Wei
: APPLICANT: Wagner, Susanne
: APPLICANT: Hess, Mark A.
: TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
: TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1
: FILE REFERENCE: Myriad 3
: CURRENT APPLICATION NUMBER: US/09/262,773
: CURRENT FILING DATE: 1999-03-04
: NUMBER OF SEQ ID NOS: 210
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 648
: TYPE: PRT
: ORGANISM: human
US-09-262-773-2

```

```

Query Match          15.6% Score 313; DB 4; Length 648;
Best Local Similarity 31.6% Pred. No. 2e-23;
Matches 95; Conservative 44; Mismatches 122; Indels 40; Gaps 17;
QY 12 SSUTLADAFIAGESSAPPPRALPRRFICSPDCSANSKAMKLDALHCKHTGERPFV 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 TNISQVNSFVNLRET---TPVHPLGRHNDGSY--CGKSTCSHLVRLHRTHTGEKPYK 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 72 CDYEGCGKAFIRDYHLNRHILHTTGKPRVCAATGCDQKFN--TKSNLKK-----HEERKH 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 C--MECGKSTYRSSHLARHOKVHKMNAFY-----KYLNRKMLEETSPVTOAERT- 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 ENQOKOYICSEFDECKTFKKHQOLKIHQOQTNELPKCTOEGCGNHFASPSKLKRNHAKA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 475 PSVEKPYRC--DDCGKHFMTSDLVNRHQRTHTEKPFCTI--CGKSFQSKSVLTTHQRI 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 HEG---YVCOKCGSFVAKTWTLLKHVRETH--KEELICEVCRKTKRKDYLRKQHKHTHA 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 HUGGKPYLIGE--CGEDFSEHRRYLAH-RKTHAAEELYLCEGCRCTHSAAFKHLRGHA 588
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 PERDVCRCPREGCGRTYTVFNLOSILSFHESRPPVCEHAGCGKTFAMKOSLTRHAYV 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 589 SVAP-CRC--NECGKSFSRDHLVRHQRTHTEKPFCTC--PTCGKSFSGYHLIRHQR 642
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 H 301
   :
Db 643 H 643

```

```

RESULT 13
US-08-486-099-117
: Sequence 117, Application US/08486099
: Patent No. 6013263
: GENERAL INFORMATION:
: APPLICANT: Bolognesi, Paul P.
: APPLICANT: Matthews, Thomas J.
: APPLICANT: Wild, Carl T.
: APPLICANT: Barney, Shawn O.
: APPLICANT: Lambert, Dennis M.
: APPLICANT: Pelletway, Stephen R.
: APPLICANT: Langlois, Alphonse J.
: TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
: TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
: NUMBER OF SEQUENCES: 209
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/486,099
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7872-031
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-9741/8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 117:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 462 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
US-08-486-099-117

```

```

Query Match          15.6% Score 312.5; DB 3; Length 462;
Best Local Similarity 32.9% Pred. No. 1.4e-23;
Matches 83; Conservative 35; Mismatches 101; Indels 33; Gaps 14;

```

[illegible]

```

RESULT 14
US-08-360-107A-127
: Sequence 127, Application US/08360107A
: Patent No. 6017536
: GENERAL INFORMATION:
: APPLICANT: Bolognesi, Daniel P.
: APPLICANT: Matthews, Thomas J.
: APPLICANT: Wild, Carl T.
: APPLICANT: Barney, Shawn O.
: APPLICANT: Lambert, Dennis M.
: APPLICANT: Petleway, Stephen R.
: APPLICANT: Langlois, Alphonse J.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
: TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV/AIDS
: TITLE OF INVENTION: TRANSMISSION
: NUMBER OF SEQUENCES: 149
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/360,107A
: FILING DATE: 20-DEC-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7872-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-9741/8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 127:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 462 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-360-107A-127
Query Match 15.6%; Score 312.5; DB 3; Length 462;
Best Local Similarity 32.9%; Pred. No. 1,4e-23;

```

	Matches	83;	Conservative	35;	Mismatches	101;	Indels	33;	Gaps	14;
QY	60	HLCKHTGSRPRVCOYDEGGCAFI	IRPYHLSRLHTLTGTEKPRFCAMTGC	COOKENTSNLKK	119					
Db	172	HOLHTGKRPVDEC	-KECGKSFSSRSHLIGKHTTGE	EPYCKE	-CGKSTSWSHLYAT	227				
QY	120	HEPKHENOQKQYICSPEDCKKTE	FKKHOOLKIHOC	HTNEPLFKCTOE	GGCGHNPASPKL	179				
Db	228	H-ORTHTG-DKLYTCN	-OCGKSPFHSSRLIHORTHTG	HKPKYECPE	-CGKSFROSTHL	281				
QY	180	KRNKAAH---	EGYVCOKGCSPFAKWTTELLK	HVR-ETNKEILLCEVC	RTKFRKIDLYKH	235				
Db	282	ILHORTHVRRVRYECNE	-CGKSYSORSHLYVNHRLHTG	ICLKPECKADCGKCF	RSRSHLYSH	340				
QY	236	MKTHAEPEDVCCPRE	-GCGRTYTVTNLSHLS	TFHEE---	SRPVECHAGCGKTA	289				
Db	341	ORTHNGEK-----	PEYCHDCGKSF	-----OSSALIVH	QRIHTGKPRYEC	-CGCGKAFI	388			
QY	290	MKOSLITRAHVVH	301							
Db	389	RKNDLIKHORIH	400							

```

RESULT 15
US-08-484-223B-117
: Sequence 117, Application US/08484223B
: Patent No. 6020459
: GENERAL INFORMATION:
: APPLICANT: Bolognesi, Danl P.
: APPLICANT: Matthews, Thomas J.
: APPLICANT: Wild, Carl T.
: APPLICANT: Barney, Shawn O.
: APPLICANT: Lambert, Dennis M.
: APPLICANT: Leteaway, Stephen R.
: TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
: TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
: TITLE OF INVENTION: TRANSMISSION
: NUMBER OF SEQUENCES: 245
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennile & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,223B
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7872-029
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-9741/8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 117:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 462 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-484-223B-117

```

Best Local Similarity 32.9%; Pred. No. 1.4e-23;
Matches 83; Conservative 35; Mismatches 101; Indels 33; Gaps 14;

[illegible]

Search completed: February 10, 2003, 17:48:58
Job time : 28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 10, 2003, 17:45:22 ; Search time 18 Seconds

(without alignments)
449.544 Million cell updates/sec

Title: US-09-831-426C-2

Perfect score: 2008

Sequence: 1 MDPAYVAESVSLTIADAF.....ESPNCVEDKMLSTAVLTIG 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 129505 seqs, 22169297 residues

Total number of hits satisfying chosen parameters: 129505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCF_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	386	19.2	406	10	US-09-764-864-976	Sequence 976, App
2	383.5	19.1	448	10	US-09-764-864-909	Sequence 909, App
3	383.5	19.1	457	10	US-09-764-864-1045	Sequence 1045, App
4	381	19.0	711	10	US-09-976-165-10	Sequence 10, Appl
5	381	19.0	711	10	US-09-828-648-2	Sequence 2, Appl1
6	379	18.9	525	10	US-09-864-761-33653	Sequence 33653, A
7	375	18.7	338	10	US-09-764-864-948	Sequence 948, App
8	372	18.5	551	10	US-09-864-761-37065	Sequence 37065, A
9	368.5	18.4	515	10	US-09-864-761-37152	Sequence 37152, A
10	366.5	18.3	278	10	US-09-864-761-42958	Sequence 42958, A
11	366.5	18.3	516	9	US-09-974-298-118	Sequence 118, App
12	366.5	18.3	547	10	US-09-764-864-1065	Sequence 1065, App
13	366	18.2	610	9	US-09-989-920-212	Sequence 212, App
14	360.5	18.0	580	10	US-09-925-301-1357	Sequence 1357, App
15	359	17.9	9	US-09-942-087A-30	Sequence 30, Appl	Sequence 18, Appl
16	359	17.9	196	10	US-09-779-223-18	Sequence 29, Appl
17	359	17.9	196	10	US-09-844-508-29	Sequence 37991, A
18	356.5	17.8	468	10	US-09-864-761-37991	Sequence 36148, A
19	356.5	17.8	536	10	US-09-864-761-36148	

20	355.5	17.7	241	10	US-09-764-864-935	Sequence 935, App
21	355.5	17.7	492	10	US-09-864-761-42897	Sequence 42897, A
22	355.5	17.7	605	10	US-09-864-761-16132	Sequence 36132, A
23	355.5	17.7	605	10	US-09-864-761-37943	Sequence 37943, A
24	354	17.6	439	10	US-09-864-761-33476	Sequence 33476, A
25	352.5	17.6	513	10	US-09-764-864-1061	Sequence 1061, App
26	351	17.5	481	10	US-09-864-761-37651	Sequence 37651, A
27	349.5	17.4	269	10	US-09-764-864-887	Sequence 887, App
28	349.5	17.4	284	10	US-09-764-864-932	Sequence 932, App
29	347.5	17.3	558	10	US-09-864-761-47724	Sequence 47724, A
30	347.5	17.3	585	10	US-09-764-864-1063	Sequence 1063, App
31	344	17.1	340	10	US-09-864-761-35894	Sequence 35894, App
32	343.5	17.1	361	10	US-09-764-864-893	Sequence 893, App
33	342.5	17.1	483	10	US-09-764-864-940	Sequence 940, App
34	342.5	17.1	614	10	US-09-864-761-43107	Sequence 43107, A
35	340.5	17.0	311	10	US-09-864-761-34333	Sequence 34333, A
36	336.5	16.8	413	10	US-09-864-761-37010	Sequence 37010, App
37	334	16.6	252	10	US-09-816-669A-12	Sequence 12, App1
38	331.5	16.5	333	10	US-09-764-864-958	Sequence 958, App
39	331	16.5	577	10	US-09-764-864-883	Sequence 883, App
40	330.5	16.5	257	10	US-09-764-864-1428	Sequence 1428, App
41	327.5	16.3	254	10	US-09-864-761-33477	Sequence 33477, A
42	326	16.2	375	10	US-09-864-761-36336	Sequence 36336, A
43	325.5	16.2	281	10	US-09-764-864-1076	Sequence 1076, App
44	324	16.1	420	10	US-09-864-761-37994	Sequence 37994, A
45	323.5	16.1	485	10	US-09-864-761-37026	Sequence 37026, A

ALIGNMENTS

RESULT 1
US-09-764-864-976
; Sequence 976, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 976
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-976

Query Match	19.2%	Score 386;	DB 10;	Length 406;
Best Local Similarity	33.9%	Pred. No. 9.8e-25;		
Matches 105;	Conservative 38;	Mismatches 125;	Indels 42;	Gaps 17;
OY	47	CSANYSKAWKLDANLHKNGEPPVCDYEGCGRAFLRDYHLSRHILPTGCKPFCVCAATG	106	
DB	82	CSMNSHLIWPQK-----HTGCKRYEC--RECGRKAFSEKSLRKHNRHTNGEKRYKC--DG	133	
OY	107	CDCKPNTKSLMKHFKKNEKNOQKUYICSPEDCKKTPFKKHOQJKNOCSTNPERLKCTQ	166	
DB	134	CDKAFKSKSLGRIN--ORTHG--EKREPC--HECGSKNYKSLIIVIORHTGTGCKPREPCNE	189	
OY	167	EGGKHKHASPASKIKRYAKNAHEG---YVCGKSGFVAKTWELKHYR--ETNKEEILCEVC	222	
DB	190	--CGKSFMSHSGILRNHRITHTGERPYKCDK--CGKAFKLGKLNKHNRTHTGERPYKCNOC	246	
OY	223	RKTFKKRDYLOKMKHTHAREPVCRCPRGCGRTYTVFVFNLOSHLSFHESSRPVCEHA	282	
DB	247	GRAPGQSQSLRGHNRHTGCKRY--KC--NHGGEAFSQSKSLVGNHNR--HTGCKRYDCEE--	301	
OY	283	GGCKTRFMKOSLGRHNAVVDPRDKKKMKLKYKKS--REKRSIASLHSGYTPPKKQGGSL	341	
DB	302	-CGKTFKROKSLNRGHORTHTGCKRYECNECGKAFSEKSVLRKH-----QR	345	

98 KPEVCATGCDQKENTKSNLKKHIEPKHENQKQYICSFEDCKTFKKHQQLKHQCQHT 157

```
Db 501 KPEYC--SDCGTFTQKSHLNH--OKIHTEGHNHVS--ECGKAFNOKSILSMHQRIHT 554
158 NEPLFCCTGEGCGKHAPSKLKHAKAHG---YVCQCGCSVATVTELLKH-VRETH 213
Db 555 GEPKYKCS--CGKATSKSQFEKHORIHTGKRPVCTE--CGKAFNGRSNFHNQIHTHR 611
Qy 214 KEELICEVCRKTEFKRDYLYKQHKHTHAPERDVCRCPRE--GCGRTYTYFNLQSHLISFH 271
Db 612 ERFPVYCKGKAFVQKSELITHQRTMGER-----PYECLDCKSKSPKQLKVH-QRIH 665
Qy 272 EESRPVCEHAGCGKTFAMKQSTFTHAVVHDPDK-KMKLYK 313
Db 666 TGERPVCSE--CGKAFNNRNSNFNKHQTHTTRDKSYKCSYSVK 706

RESULT 5
US-09-828-648-2
; Sequence 2, Application US/09828648
; Patent No. US20020151510A1
; GENERAL INFORMATION:
; APPLICANT: Ikezu, Tsuneya
; APPLICANT: Leisman, Gary
; APPLICANT: Carlson, Kimberly A.
; APPLICANT: Gendelman, Howard E.
; APPLICANT: University of Nebraska Medical Center
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treatment of Human Immunodeficiency Virus Infection
; FILE REFERENCE: UNMK 63156
; CURRENT APPLICATION NUMBER: US/09/828,648
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/246,331
; PRIOR FILING DATE: 2000-11-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 711
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-828-648-2

Query Match 19.0%; Score 381; DB 10; Length 711;
Best Local Similarity 32.9%; Pred. No. 5e-24;
Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;

Qy 38 RRTICSPPOCSANYSKAMLDALCHTGERPVCDEGCGKAFIDYHLSRIILHTGCE 97
Db 445 KSYVCI--ECGAFIQKAMHLIVHQRSHTEGKPYQC--HMGCKSFISKSQLDIHNRHTGE 500
Qy 98 KPEVCATGCDQKFNKTNLKHFERKHNENQOYICSPEDCKKTFKKHQLKHOCQHT 157
Db 501 KPEYC--SDCGTFTQKSHLNH--OKIHTEGHNHVS--ECGKAFNOKSILSMHQRIHT 554
Qy 158 NEPLFCCTGEGCGKHAPSKLKHAKAHG---YVCQCGCSVATVTELLKH-VRETH 213
Db 555 GEPKYKCS--CGKATSKSQFEKHORIHTGKRPVCTE--CGKAFNGRSNFHNQIHTHR 611
Qy 214 KEELICEVCRKTEFKRDYLYKQHKHTHAPERDVCRCPRE--GCGRTYTYFNLQSHLISFH 271
Db 612 ERFPVYCKGKAFVQKSELITHQRTMGER-----PYECLDCKSKSPKQLKVH-QRIH 665
Qy 272 EESRPVCEHAGCGKTFAMKQSTFTHAVVHDPDK-KMKLYK 313
Db 666 TGERPVCSE--CGKAFNNRNSNFNKHQTHTTRDKSYKCSYSVK 706

RESULT 6
US-09-864-761-33653
; Sequence 33653, Application US/09864761
; Patent No. US200200446763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
```

```
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmice-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33653
; LENGTH: 525
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 284476.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBLL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-106
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-106
US-09-864-761-33653

Query Match 18.9%; Score 379; DB 10; Length 525;
Best Local Similarity 32.7%; Pred. No. 5.1e-24;
Matches 101; Conservative 39; Mismatches 123; Indels 46; Gaps 13;

Qy 47 CSANYSKAMLDALCHTGERPVCDEGCGKAFIDYHLSRIILHTGKRPVCATG 106
Db 111 CGNFFWSHSDLIHEDQIHSEKRPVQCN--ECGKAFTRNQLSMHRIILHTGKRPNC--TQ 166
```

```
QY 107 CQCKENTKSNLKKHFKRKHENQOQYICSFEDCKTKFKKHQOLKHQCSHTNPLEKCTQ 166
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 167 CGKAFNSRSLGRH--KTHSGEKPHEC--RDCGKAFKTRNRLCMQLHTGKRPKC-- 220
QY 167 ECGKHFAPSKLRHAKHNEG--YVCOKGSFVAKTWTLLKHR-ETHKEILLCEVC 222
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 221 NCCGKAFQKHSHTTIGRLHTGKPECEB-CGKAFSGSSDLTKHRLHTGEPYCSKC 279
QY 223 RKTFRKDYLKQMKHTAERDVCRRPGCGRTYTVFNLSHILSFHEESRPFVCEHA 282
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 280 GKAFFSSSLSKKRIHTREKH-GCPO--CGKDFSIKMLTGH-RRIHTEEKRYGCEEC 335
QY 283 G-----CGKTFAMKOSLTRHAAVVDPRDKMKMLKYKKS- 315
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 336 GKAFFRNCRAHREHRTGEKPYOCDRCGKTFODKHCLTIHORHTGKRPYKCLECGKAF 395
QY 316 REKRSLASH 324
      | : | : |
Db 396 SGKSNLTGH 404
```

```
RESULT 7
US-09-764-864-948
; Sequence 948, Application US/09764864
; Patent No. US2002013273A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIORITY FILING DATE: 2001-01-17
; PRIORITY APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 948
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (335)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-948
```

```
Query Match          18.7%: Score 375; DB 10; Length 338;
Best Local Similarity 31.5%: Pred. No. 6, 5e-24;
Matches 93; Conservative 49; Mismatches 125; Indels 28; Gaps 14;
```

```
QY 38 RPFICSFPPCSANYSKAMKLDHNLCKHTEGRPVCDYEGCGKAFIDYHLSRHILTHTGE 97
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 12 KRWQCMH--CEKSYTSKYNLVTHILGNHGIKPHSCPH--CSKLFKOPSHLOTHTLLTHOCT 67
QY 98 KPVCAATGCDQKFNKSNLKKHFKRKHENQOQYICSFEDCKTKFKKHQOLKHQCSHT 157
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 68 RPKKCOV--CHKAFYOTSHIKRIM--LLHSEVKPYSCHE--GORGAYVSELANHEVKNIE 121
QY 158 NEPLFECTGEGGKHFAPSKLRHAKHNEG-----YVCOKGSFVAKTWTLLKHYVE 211
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 122 SCGRCHVCE--CGLDRESLTLQKRLHSLHSGPTLYQCLSDCKSFHYRSDQONMLKH-- 176
QY 212 THKEILLCEVCRTFKRKDYLKQMKHTAERDVCRRPGCGRTYTVFNLSHILSLHSLH 271
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 177 QNVRPVCTECGMEFSQIHNLKQHSILTHGKVEF--KC--EVCGRFPTLDANMKRML-TH 232
QY 272 EESRPVCEHAGCGTFAKOSLTRHAAVVDPRDK-KMKLKVKKSEKKSLSLASHL 325
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 233 TVSRPQOC-HI-CEKTFVOKOTLTKHTMIVHSPKPCVCGKSFNRMYNLTGHM 285
```

```
RESULT 8
US-09-864-761-37065
; Sequence 37065, Application US/09864761
; Patent No. US20020048763A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aomicla-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIORITY FILING DATE: 2001-05-23
; PRIORITY APPLICATION NUMBER: US 60/180,312
; PRIORITY FILING DATE: 2000-02-04
; PRIORITY APPLICATION NUMBER: US 60/207,456
; PRIORITY FILING DATE: 2000-05-26
; PRIORITY APPLICATION NUMBER: US 09/632,366
; PRIORITY FILING DATE: 2000-08-03
; PRIORITY APPLICATION NUMBER: GB 24263.6
; PRIORITY FILING DATE: 2000-10-04
; PRIORITY APPLICATION NUMBER: US 60/236,359
; PRIORITY FILING DATE: 2000-09-27
; PRIORITY APPLICATION NUMBER: PCT/US01/00666
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00667
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00664
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00669
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00665
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00668
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00663
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00662
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00661
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: PCT/US01/00670
; PRIORITY FILING DATE: 2001-01-30
; PRIORITY APPLICATION NUMBER: US 60/234,687
; PRIORITY FILING DATE: 2000-09-21
; PRIORITY APPLICATION NUMBER: US 09/608,408
; PRIORITY FILING DATE: 2000-06-30
; PRIORITY APPLICATION NUMBER: US 09/774,203
; PRIORITY FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37065
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; MAP TO AC005261.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: SWISSPROT HIT: P51522, EVALUATE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-105
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-105
US-09-864-761-37065
```

```
Query Match          18.5%: Score 372; DB 10; Length 551;
Best Local Similarity 30.2%: Pred. No. 2, 1e-23;
Matches 106; Conservative 45; Mismatches 124; Indels 76; Gaps 18;
```



```

Db      235  EKPECKE--CGKAFGLTOLAHOHIHTGKC 264

RESULT 11
US-09-974-298-118
; Sequence 118: Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: 60/236,331
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO: 118
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 5665139CD1
US-09-974-298-118

Query Match      18.3%; Score 366.5; DB 9; Length 516;
Best Local Similarity 35.3%; Pred. No. 5,5e+23;
Matches 94; Conservative 36; Mismatches 105; Indels 27; Gaps 14;

Oy      47  CSANYSKAMKLDALHLCIKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGKRPFYCATG 106
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      238  CGKVSRSNSYLVQHLITHTGKRPYCNV--CGKVFHHSIHLAHOIRHTGKRPYKC--NE 293

```

```

RESULT 12
US-09-764-864-1065
? Sequence 1065, Application US/09764864
? Patent No. US20020132753A1
? GENERAL INFORMATION:
? APPLICANT: Rosen et al.
? TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
? FILE REFERENCE: PT223
? CURRENT APPLICATION NUMBER: US/09/764,864
? CURRENT FILING DATE: 2001-01-17
? Prior application data removed - consult PALM or file wrapper
? NUMBER OF SEQ ID NOS: 1792
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1065
? LENGTH: 547
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (289)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE

```

LOCATION: (290)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1065

Query Match 18.3%; Score 366.5; DB 10; Length 547;
Best Local Similarity 29.9%; Pred. No. 5,9e-23;

Matches 104; Conservative 47; Mismatches 116; Indels 81; Gaps 18;

QY 38 RRFICSPDCSANYSKAMLDANLCKHTGERPVCYEGCGKAFITDYLSRIILHTGCE 97
DB 214 RHTICS--HCGKAFSISDLNRHOKTHTGDRPYKC-YE-CGKGFSRSSHLIOHRTHTGCE 269
QY 98 KPFVCAATG-----CDOKNTKSNLKKHFERKHENOQKQY 132
DB 270 RPYDCNECKSGRGRSHLIXXNHPTEGPKHCKNECKSFICRLSHLLOH-QRTHTSG-EKPY 327
QY 133 ICSFEDCKKTFKKHKOOLKTHOCOTNEPLFKCTOEGCGKHNFPASPKLKRHAANEH--Y 189
DB 328 EC--EECGKSFSSSHLAOHORTHTGCEKPEECNE--CGRGSERSDLIKHYVHTGERPY 383
QY 190 VCGKGSFPAKWTTELKTHVR-ETHKEELCEVCGRKTFKRDVLYKQNMKTNAERDVCR 248
DB 384 KODE-CGKNFSONSDLVRIHRAHTGKPYHCNCGENFESRISHLVOHRTHTGEC-----HAG-- 437
QY 249 PRE--CGGRYTYTVFMLOSHLS--FHESRPFVCE-----HAG-- 283
DB 438 PYECNACGKS---FSRSHLITHTOKIHTGKPYECNECKMRSFGERSDLIKHRTHTGEC 493
QY 284 -----CGTFAMKOSLTRHVAVHDPDKKMKLKYKRSRKS-LASH 324
DB 494 PYECVCGKGFTOSNLTHTORVHTGKPEYECTECKSFSRSALIKH 541

RESULT 13

US-09-989-920-212
Sequence 212, Application US/09989920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods relating to Lung Specific Genes and Prot
FILE REFERENCE: DEX-0291
CURRENT FILING DATE: 2001-11-21
PRIOR FILING DATE: 2000-11-22
PRIOR APPLICATION NUMBER: 60/252,500
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 212
LENGTH: 610
TYPE: PRT
ORGANISM: Homo sapien
US-09-989-920-212

Query Match 18.2%; Score 366; DB 9; Length 610;
Best Local Similarity 27.1%; Pred. No. 7,4e-23;

Matches 116; Conservative 52; Mismatches 136; Indels 124; Gaps 20;

QY 10 SVSLTIADAFIAGSSAPTPRPALPRFICSPDCSANYSKAMLDANLCKHTGERP 69
DB 117 SYNSLLDHRHTHTGE-----KRYVDV--CGTFNNAGLKYHRLHTGECR 162
QY 70 FVCDYEGCGKAFITDYLSRIILHTGKRPVCAATGCDOKNTKSNLKKHFERKHENOQ 129
DB 163 YKCDV--CGKAYISRSSLKHGKIHGKPYKSY--CEKSFYSSALDQH--KRIHTRE 216
QY 130 KQYICFEDCKKTFKKHKOOLKTHOCOTNEPLFKCTOEGCGKHNFPASPKLKRHAANEH 189
DB 217 KPEGC--DECGKAFRNNSSGLKHYKRIHTGTERPYKC--EECGKAYISLSSLINHKSVPGE 272

QY 190 -----VCGKG-----CSFYAKT--WTELLKHYRETKEE--IL 218
DB 273 KPECKDECKAFITTYRTLTNNKKVHLGKERYKCDVCEKSFNTSLSQHRRVHTREKPYE 332
QY 219 CEVCGRTEFRKRDVLYKQNMKTNAER---DVC-----RCPR--EGC 253
DB 333 CDRCEKVFRRNSSLKHYKRIHTGERPERYECDDVCGKAYISHSLINHKSTHGPRTHTGCE 392
QY 254 GRITTYTVFMLOSH-----ILSFHEE---SRPFVCAAGCGK 286
DB 393 GKAFSSRSLISKHYVHLGKRPVCECKSFSYSSLSQHKRIHTGKRPVYCDR--CGK 450
QY 287 TFAMKOSLTRHVAVHDPDK-----KKMKLYKRSREKRSILASHLSGYPKRGOGGLSD 341
DB 451 APRNSSLGLVTHKRIHTGKPYECDECGKAYISHS--SLINKSVH-----OGKOPYN 500
QY 342 CONGESPN 349
DB 501 CEGCKSFN 508

RESULT 14

US-09-925-301-1357
Sequence 1357, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1357
LENGTH: 580
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (526)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1357

Query Match 18.0%; Score 360.5; DB 10; Length 580;
Best Local Similarity 34.7%; Pred. No. 2e-22;

Matches 92; Conservative 32; Mismatches 118; Indels 23; Gaps 13;

QY 46 DCSANSKAMLDANLCKHTGERPVCYEGCGKAFITDYLSRIILHTGKRPVCAAT 105
DB 247 ECGKGRYSAYVLIHMGRTHTGDMPTYC--KECGKAFTRSCQLTQHKRTHTGKRPVCK--K 302
QY 106 GCDOKNTKSNLKKHFERKHENOQKQYICSFEDCKKTFKKHKOOLKTHOCOTNEPLFKCT 165
DB 303 DCGRAFTVSSLSQHM--KIHVEKPEYC--KECGIAFTRSSQLTEHLKHTAKDPFEC-- 357
QY 166 OEGCGKHNFPASPKLKRHAANEH---YVCGKGSFPAKWTTELKTHVRHTKEE--ILCE 220
DB 358 -KICGKSFRRNSCLSDHFRHTGTIKPYKC-KDCGKAFTONSDLTQHAR-THSERPYECK 414
QY 221 VCRKTRKRDVLYKQNMKTNAERDVCRCPREGGRTTYTVFMLOSHLSFHESSRPFVCE 280
DB 415 ECGKAFARSSRLSEHRTHTGKRPF--ECYK--CGKAFAISSNLGSH-LRIHTGKRPPEC-- 469
QY 281 HAGCGKTFAMKOSLTRHVAVHDPDK 305
DB 470 -LECGKAFTHSSSLNNHMRTHSAKK 493

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:47:38 ; Search time 300 seconds
(without alignments)
784.425 Million cell updates/sec

Title: US-09-831-426c-2

Perfect score: 2008
Sequence: 1 MDPFAVAESVSLTIADAF.....ESPNCVEDKMLSTVAVLFIG 365

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/1/paa/PCFUS_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	100.0	365	22	US-09-831-426c-2
2	2008	100.0	365	22	US-09-831-426c-2
3	2000	99.6	409	1	PCT-US00-26524B-5945
4	2000	99.6	409	25	US-10-106-698-5955
5	821	40.9	184	1	PCT-US01-18569-2978
6	821	40.9	184	26	US-10-264-049-2978

7	556	27.7	106	21	US-09-760-466-1024	Sequence 1024, Ap
8	556	27.7	106	21	US-09-760-466-1024	Sequence 1024, Ap
9	556	27.7	106	26	US-10-212-083-1024	Sequence 255, Ap
10	556	27.7	106	26	US-10-223-026-255	Sequence 255, Ap
11	453	22.6	95	1	PCT-US01-08631-39706	Sequence 39706, A
12	425	21.2	927	1	PCT-US01-04098A-3929	Sequence 3929, Ap
13	421	21.0	869	1	PCT-US01-04098A-1961	Sequence 1961, Ap
14	417.5	20.8	803	1	PCT-US01-04098A-1676	Sequence 1676, Ap
15	417.5	20.8	803	24	US-10-029-386-32185	Sequence 32185, A
16	417.5	20.8	803	27	US-60-388-967-1355	Sequence 1355, Ap
17	417.5	20.8	803	27	US-60-412-418-1355	Sequence 1355, Ap
18	414.5	20.6	904	1	PCT-US01-04098A-3644	Sequence 3644, Ap
19	410.5	20.4	499	25	US-10-108-260A-2954	Sequence 2954, Ap
20	410.5	20.4	553	1	PCT-US02-30474-1389	Sequence 1389, Ap
21	410.5	20.4	553	27	US-60-324-631-1394	Sequence 1394, Ap
22	410.5	20.4	622	1	PCT-US01-04098A-1609	Sequence 1609, Ap
23	410.5	20.4	631	1	PCT-US01-04098A-3577	Sequence 3577, Ap
24	410.5	20.4	632	1	PCT-US01-08631-48745	Sequence 48745, A
25	410.5	20.4	648	1	PCT-US02-21179-32	Sequence 32, Appl
26	410.5	20.4	652	1	PCT-US01-04098A-1833	Sequence 1833, Ap
27	410	20.4	363	27	US-60-388-055-14	Sequence 14, Appl
28	410	20.4	434	1	PCT-US02-28540-6	Sequence 6, Appl
29	409	20.4	620	1	PCT-US02-28540-54	Sequence 54, Appl
30	406.5	20.2	403	20	US-09-614-150-20385	Sequence 20385, A
31	406.5	20.2	403	27	US-60-167-217-20542	Sequence 20542, A
32	406.5	20.2	403	27	US-60-173-464-16736	Sequence 16736, A
33	406.5	20.2	403	27	US-60-191-637-20447	Sequence 20447, A
34	406.5	20.2	403	27	US-60-191-681-16100	Sequence 16100, A
35	406.5	20.2	587	1	PCT-US00-27620-49	Sequence 49, Appl
36	406.5	20.2	700	18	US-09-488-725A-3146	Sequence 3146, Ap
37	406	20.2	693	27	US-60-117-905-56	Sequence 56, Appl
38	405	20.2	816	25	US-10-104-047-2341	Sequence 2341, Ap
39	405	20.2	1241	1	PCT-US01-08631-33156	Sequence 33156, A
40	404	20.1	779	1	PCT-US01-04098A-1636	Sequence 1636, Ap
41	404	20.1	779	1	PCT-US01-04098A-3604	Sequence 3604, Ap
42	404	20.1	779	1	PCT-US01-08631-31239	Sequence 31239, A
43	403.5	20.1	365	27	US-60-388-055-16	Sequence 16, Appl
44	403	20.1	365	27	US-60-388-055-11	Sequence 11, Appl
45	403	20.1	487	24	US-10-029-386-33849	Sequence 33849, A

ALIGNMENTS

RESULT 1
US-09-831-426c-2
; Sequence 2, Application US/09831426
; GENERAL INFORMATION:
; APPLICANT: Hoechst Marion Roussel
; APPLICANT: Bordon-Pallier, F.
; APPLICANT: Rocher, C.
; TITLE OF INVENTION: Human htfIIIA gene and coded htfIIIA protein
; FILE REFERENCE: 146.1364
; CURRENT APPLICATION NUMBER: US/09/831,426
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 2
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Human
US-09-831-426c-2

Query Match 100.0%; Score 2008; DB 22; Length 365;
Best Local Similarity 100.0%; Pred. No. 4,8e+164;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPFAVAESVSLTIADAFIAGESSAPPPRALPRRIFCSFPDCSANYSKAMKIDAH 60
DB 1 MDPFAVAESVSLTIADAFIAGESSAPPPRALPRRIFCSFPDCSANYSKAMKIDAH 60
QY 61 LCKHTGERPVCDCYEGCGKAFIRDYHLSRIILHTGTGKPVCAATGCDGKFNKSNLKKH 120

```
Db 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
QY 121 FERKHNQOQOYICSPEDCKKTFKKHOOLKIHOCQHTNEPLFCSTOEGCGKHPASPSKLK 180
Db 121 FERKHNQOQOYICSPEDCKKTFKKHOOLKIHOCQHTNEPLFCSTOEGCGKHPASPSKLK 180
QY 181 RHAKAHGVCYQKGCSEFVAKTWTLLKHNRETHKEELICEVCKTKTKRKDYLKQHNKHTHA 240
Db 181 RHAKAHGVCYQKGCSEFVAKTWTLLKHNRETHKEELICEVCKTKTKRKDYLKQHNKHTHA 240
QY 241 PERDVCRCPREGCGRTTYTFVFNLOSHILSFHESRPVCEHACCGKTFAMKOSLTRHAYV 300
Db 241 PERDVCRCPREGCGRTTYTFVFNLOSHILSFHESRPVCEHACCGKTFAMKOSLTRHAYV 300
QY 301 HDPDKKMKMLKVKKSREKRSLSHLSGYIPPKKOGGSLCONGESPNCVEDKMLSTVA 360
Db 301 HDPDKKMKMLKVKKSREKRSLSHLSGYIPPKKOGGSLCONGESPNCVEDKMLSTVA 360
QY 361 VLTLG 365
Db 361 VLTLG 365

RESULT 2
US-09-831-426c-2
; Sequence 2, Application US/09831426C
; GENERAL INFORMATION:
; APPLICANT: Hoechst Marion Roussel
; APPLICANT: Borden-Pallier, F.
; APPLICANT: Rocher, G.
; TITLE OF INVENTION: Human htfIIIA gene and coded htfIIIA protein
; FILE REFERENCE: 146.1364
; CURRENT APPLICATION NUMBER: US/09/831,426C
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 2
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Human
US-09-831-426c-2

Query Match 100.0%; Score 2008; DB 22; Length 365;
Best Local Similarity 100.0%; Pred. No. 4, 8e-164;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPRAVAESVSLTTADAFIAGESSAPTPRRPALPRRFICSPDCSANYSKAMKLDH 60
Db 1 MDPRAVAESVSLTTADAFIAGESSAPTPRRPALPRRFICSPDCSANYSKAMKLDH 60
QY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
Db 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
QY 121 FERKHNQOQOYICSPEDCKKTFKKHOOLKIHOCQHTNEPLFCSTOEGCGKHPASPSKLK 180
Db 121 FERKHNQOQOYICSPEDCKKTFKKHOOLKIHOCQHTNEPLFCSTOEGCGKHPASPSKLK 180
QY 181 RHAKAHGVCYQKGCSEFVAKTWTLLKHNRETHKEELICEVCKTKTKRKDYLKQHNKHTHA 240
Db 181 RHAKAHGVCYQKGCSEFVAKTWTLLKHNRETHKEELICEVCKTKTKRKDYLKQHNKHTHA 240
QY 241 PERDVCRCPREGCGRTTYTFVFNLOSHILSFHESRPVCEHACCGKTFAMKOSLTRHAYV 300
Db 241 PERDVCRCPREGCGRTTYTFVFNLOSHILSFHESRPVCEHACCGKTFAMKOSLTRHAYV 300
QY 301 HDPDKKMKMLKVKKSREKRSLSHLSGYIPPKKOGGSLCONGESPNCVEDKMLSTVA 360
Db 301 HDPDKKMKMLKVKKSREKRSLSHLSGYIPPKKOGGSLCONGESPNCVEDKMLSTVA 360
QY 361 VLTLG 365
Db 361 VLTLG 365
```

```
RESULT 3
PCT-US00-26524B-5945
; Sequence 5945, Application PCT/US0026524B
; GENERAL INFORMATION:
; APPLICANT: Birse et. al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005PCT
; CURRENT APPLICATION NUMBER: PCT/US00/26524B
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5945
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-26524B-5945

Query Match 99.6%; Score 2000; DB 1; Length 409;
Best Local Similarity 99.5%; Pred. No. 2, 7e-163;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPRAVAESVSLTTADAFIAGESSAPTPRRPALPRRFICSPDCSANYSKAMKLDH 60
Db 45 LDPRAVAESVSLTTADAFIAGESSAPTPRRPALPRRFICSPDCSANYSKAMKLDH 104
QY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 120
Db 105 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILHTGEEKPFVCAATGCDOKFNTKSNLKKH 164
QY 121 FERKHNQOQOYICSPEDCKKTFKKHOOLKIHOCQHTNEPLFCSTOEGCGKHPASPSKLK 180
Db 165 FERKHNQOQOYICSPEDCKKTFKKHOOLKIHOCQHTNEPLFCSTOEGCGKHPASPSKLK 224
QY 181 RHAKAHGVCYQKGCSEFVAKTWTLLKHNRETHKEELICEVCKTKTKRKDYLKQHNKHTHA 240
Db 225 RHAKAHGVCYQKGCSEFVAKTWTLLKHNRETHKEELICEVCKTKTKRKDYLKQHNKHTHA 284
QY 241 PERDVCRCPREGCGRTTYTFVFNLOSHILSFHESRPVCEHACCGKTFAMKOSLTRHAYV 300
Db 285 PERDVCRCPREGCGRTTYTFVFNLOSHILSFHESRPVCEHACCGKTFAMKOSLTRHAYV 344
QY 301 HDPDKKMKMLKVKKSREKRSLSHLSGYIPPKKOGGSLCONGESPNCVEDKMLSTVA 360
Db 345 HDPDKKMKMLKVKKSREKRSLSHLSGYIPPKKOGGSLCONGESPNCVEDKMLSTVA 404
QY 361 VLTLG 365
Db 405 VLTLG 409

RESULT 4
US-10-106-698-5955
; Sequence 5955, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
```



```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and Antibodies
; FILE REFERENCE: PT259
; CURRENT APPLICATION NUMBER: US/09/760,466
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1813
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 1024
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-760-466-1024

Query Match          27.7%: Score 556; DB 21; Length 106;
Best Local Similarity 100.0%; Pred. No. 1,3e-39;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 VFNLSHLSFHEESRPFYCEHAGCGKTFAMKQSLTRHVVHDPDKKKKLKVKKSREKR 319
Db 1 VFNLSHLSFHEESRPFYCEHAGCGKTFAMKQSLTRHVVHDPDKKKKLKVKKSREKR 60

Qy 320 SLASHLSGYIPPKRKGGLSLCQNGESPNCVEDKMLSTVAVLTIG 365
Db 61 SLASHLSGYIPPKRKGGLSLCQNGESPNCVEDKMLSTVAVLTIG 106

RESULT 8
US-09-760-494-255
; Sequence 255, Application US/09760494
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and Antibodies
; FILE REFERENCE: PC018
; CURRENT APPLICATION NUMBER: US/09/760,494
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 255
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-760-494-255

Query Match          27.7%: Score 556; DB 21; Length 106;
Best Local Similarity 100.0%; Pred. No. 1,3e-39;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 VFNLSHLSFHEESRPFYCEHAGCGKTFAMKQSLTRHVVHDPDKKKKLKVKKSREKR 319
Db 1 VFNLSHLSFHEESRPFYCEHAGCGKTFAMKQSLTRHVVHDPDKKKKLKVKKSREKR 60

Qy 320 SLASHLSGYIPPKRKGGLSLCQNGESPNCVEDKMLSTVAVLTIG 365
Db 61 SLASHLSGYIPPKRKGGLSLCQNGESPNCVEDKMLSTVAVLTIG 106

RESULT 9
US-10-212-083-1024
; Sequence 1024, Application US/10212083
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT259CIN
; CURRENT APPLICATION NUMBER: US/10/212,083
; PRIOR FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: 09/760,466
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
```

```

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1813
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 1024
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-212-083-1024

Query Match          27.7%: Score 556; DB 26; Length 106;
Best Local Similarity 100.0%; Pred. No. 1,3e-39;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 VFNLSHLSFHEESRPFYCEHAGCGKTFAMKQSLTRHVVHDPDKKKKLKVKKSREKR 319
Db 1 VFNLSHLSFHEESRPFYCEHAGCGKTFAMKQSLTRHVVHDPDKKKKLKVKKSREKR 60

Qy 320 SLASHLSGYIPPKRKGGLSLCQNGESPNCVEDKMLSTVAVLTIG 365
Db 61 SLASHLSGYIPPKRKGGLSLCQNGESPNCVEDKMLSTVAVLTIG 106

RESULT 10
US-10-223-026-255
; Sequence 255, Application US/10223026
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC018CIN
; CURRENT APPLICATION NUMBER: US/10/223,026
; PRIOR FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: 09/760,494
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 255
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
```


US-10-223-026-255

Query Match 27.7%: Score 556; DB 26; Length 106;
Best Local Similarity 100.0%; Pred. No. 1,3e-39;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 VFNLSHLSFHEESRPVCEHAGCGKTFAMKQSLTRHAVVHDPDKKKMLKYKKSREKR 319
DB 1 VFNLSHLSFHEESRPVCEHAGCGKTFAMKQSLTRHAVVHDPDKKKMLKYKKSREKR 60
QY 320 SLASHLSGYTPPKRKGGLSLCNGESPNCVEDKMLSTVAVLTIG 365
DB 61 SLASHLSGYTPPKRKGGLSLCNGESPNCVEDKMLSTVAVLTIG 106

RESULT 11

PCT-US01-08631-39706
Sequence 39706, Application PC/TUS0108631

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 21272-049

CURRENT APPLICATION NUMBER: PCT/US01/08631

CURRENT FILING DATE: 2001-03-30

PRIOR APPLICATION NUMBER: 09/540,217

PRIOR FILING DATE: 2000-03-31

PRIOR APPLICATION NUMBER: 09/649,167

PRIOR FILING DATE: 2000-08-23

NUMBER OF SEQ ID NOS: 60736

SOFTWARE: Custom

SEQ ID NO 39706

LENGTH: 95

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: DOMAIN

LOCATION: (8)..(22)

OTHER INFORMATION: C2H2-TYPE ZINC FINGER SIGNATURE domain identified by EMATRIX.

PCT-US01-08631-39706

Query Match 22.6%: Score 453; DB 1; Length 95;
Best Local Similarity 91.6%; Pred. No. 8,6e-31;
Matches 87; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 271 HESRPVCEHAGCGKTFAMKQSLTRHAVVHDPDKKKMLKYKKSREKSLASHLSGYTP 330
DB 1 HESRPVCEHAGCGKTFAMKQSLTRHAVVHDPDKKKMLKYKKSREKSLASHLSGYTP 60

QY 331 PKRKGGLSLCNGESPNCVEDKMLSTVAVLTIG 365
DB 61 PKRKGGLSLCNGESPNCVEDKMLSTVAVLTIG 95

RESULT 12

PCT-US01-04098A-3929

Sequence 3929, Application PC/TUS0104098A

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides

FILE REFERENCE: 21272-029

CURRENT APPLICATION NUMBER: PCT/US01/04098A

CURRENT FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: Not Yet Assigned

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: 09/728,422

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 09/693,325

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 09/663,561

PRIOR FILING DATE: 2000-09-15

PRIOR APPLICATION NUMBER: 09/654,936

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 3929
LENGTH: 927
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-04098A-3929

Query Match 21.2%: Score 425; DB 1; Length 927;
Best Local Similarity 32.0%; Pred. No. 4,2e-27;
Matches 116; Conservative 42; Mismatches 115; Indels 89; Gaps 19;

QY 3 PPAVVAESVSLTADAFIAGESSAPTPRPA-LRRFTCSFPDGSANYSKAKIDAH 61
DB 183 PPAVVAESVSLTADAFIAGESSAPTPRPA-LRRFTCSFPDGSANYSKAKIDAH 227
QY 62 CKHTGSRPFVCDYEGCGKAFTRDYHLSRHILTHGKPEVC----- 102
DB 228 RHHTGERPNTG--SECGKSTQSHLVQHORHTGKPEYKCPGCGKCFSSMLVQHORT 285
QY 103 -----ATGCDQKFNYSNLKHKFERKHENOQOYICSEFEDCKTFKKHQOLKIHQO 155
DB 286 HTGKPEYKCTECHKAFQTSQNLKHK-ORSHTG-EKPEYKCG--ECRAFAVSSOLIOHQA 341
QY 156 HTNEPLPKCTOEGCGKHFAPSKLKHAKHAG---YVQK-GCSFV-----AKTPT 203
DB 342 HTGKPEYKCPG--CGKRFQGNHNLKHOKIHAGEKPRCTECKSFTQSELTQHOFHT 399
QY 204 -----ELKLVHRETTHKEE--ILCEVCKTFPRRDYLNKOHKHTAPER 243
DB 400 GEKPEYCELECGKSFQHSSTLIKHKR--THLREDFPKCPVCGKTTLSATLRLHORTTGER 458
QY 244 DVCRCRECGGRTYTVFNLSHLSFHEESRPVCEHAGCGKTFAMKQSLTRHAVVHDP 303
DB 459 PV-KCPE--CGKSFVSSNLINH-ORIHGERPYIC--ADCGKSFIMSSFTLIRHORIHTG 512
QY 304 DK 305
DB 513 EK 514

RESULT 13
PCT-US01-04098A-1961
Sequence 1961, Application PC/TUS0104098A

GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-029
CURRENT APPLICATION NUMBER: PCT/US01/04098A
CURRENT FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: Not Yet Assigned
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875

```

; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1961
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1961

```

```

Query Match      21.0%; Score 421; DB 1; Length 869;
Best Local Similarity 32.0%; Pred. No. 8,6e-27;
Matches 116; Conservative 41; Mismatches 116; Indels 89; Gaps 19;

```

```

QY 3 PAVVAVESLSLIAAFIAAGESSAPTPPPRA-LPRRFICSPDDCSANYSKAMKIDAH 61
DB 125 PAVVAVNEPSLRELVG-----RPAAGAKPYICN--ECGKSPSOWSKLLRHQ 169
QY 62 CKHTGCRPVCDYEGCGKAFIRDYHLNRHLTHTGKRPVC-----102
DB 170 RHHTGRRPTC--SECGKFTQSHLVQHRTHTGKRPYKCRPCCGKCFMSNLYVQHRT 227
QY 103 -----AATGCDQKNTKSNLKKHFERKHENOQKOYICSPEDCKTKFKHQOLKHQO 155
DB 228 HNGEERPKCTEGCIAFTOSTNLIKH-QRSHTC-EKPYKCG--ECRAFAFRSSDLIOHQT 283
QY 156 HNEPELFKCTOEGCGKHAFSPSKLRHAKAHG---YVQCK-GCSFV-----AKWT 203
DB 284 HNGEERPKCTEGCIAFTOSTNLIKH-QRSHTC-EKPYKCG--ECRAFAFRSSDLIOHQT 341
QY 204 -----ELDKHRETTHKEE--ILCEVCRTKFRKDYLOKHKHTAPER 243
DB 342 GKKPYECLECGKSFHSSFLIKHQR-THLRBDPKCPVCGKFTPLSATLRLHQRTTGER 400
QY 244 DVCRCRCCGRTYTTVFNLSHILSFHEBSRPVCEHAGCGKTFAMKOSLTRHAYVDP 303
DB 401 PV-KCPE--CGKSFVSNSLNH-QRIHNGERPIC--ADCGKSFMTSSLRHORIHGT 454
QY 304 DK 305
DB 455 EK 456

```

```

RESULT 14
PCT-US01-04098A-1676

```

```

; Sequence 1676, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom

```

```

; SEQ ID NO 1676
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1676

```

```

Query Match      20.8%; Score 417.5; DB 1; Length 803;
Best Local Similarity 29.1%; Pred. No. 1,6e-26;
Matches 111; Conservative 47; Mismatches 131; Indels 93; Gaps 14;

```

```

QY 18 DAFIAAGESSAPTP-----PRALPRR-----FICSPDDCSANYSKAMKID 58
DB 230 DLIAPAPAPAPAPAEBAEAGCPAALGPRGLSGPGVLYLCPBAOCGOTFAKKHOLK 289
QY 59 AHLCKHT---GRRPVCDYEGCGKAFIRDYHLNRHLTHTGKRPVCANRCCOQKNTKS 115
DB 290 VHLTHSSGOGRRPCPLGCGCMTTSTYKLRHLQSHDKLRPFCCPABCGCKSFTTY 349
QY 116 NLKHP-----ERKHENOQKOYICSPEDCKTKFKHQO 148
DB 350 NLKANKKHGHOENSFPCVCEESFTQAKLSAHQSHFERPERFYQCAFSGCKRTFTVSA 409
QY 149 LKHOCQHTNE-PLFKCTDCCGKHAFASPSKLRHAKAHG---YVQ-KGCSFVAKWT 203
DB 410 LFSHNRAHREOELFSCSPGCSKOYDKACRLKIHLSHTGRRPFLCDRDCGGMFTSMS 469
QY 204 ELKHVRETHKEEILCEVCRTKFRKDYLOKHKHTAPERDVCRCRCCGRTYTTVFN 263
DB 470 KLRH-----HDDRRP-MCEPVEGCGSFTRAEHL 501
QY 264 OSHILFHEBSRPVCEHAGCGKTFAMKOSLTRHAYVH--DPRKKMKLV-----KSR 317
DB 502 KGHST-HLGTKRPVCGVAGCCARFASRSILYHSKHLQDVDTWMSRCPDSSCNLTFS 560
QY 318 KSLASHLSGIYPPKRRQOGL 339
DB 561 KHSMTM---VRRKVGODL 578

```

```

RESULT 15

```

```

; Sequence 32185, Application US/10029386
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hankel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32185
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 299130.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: SWISSPROT HIT: P98169, EVALU0 0.00e+00
US-10-029-386-32185

```

```

Query Match      20.8%; Score 417.5; DB 24; Length 803;
Best Local Similarity 29.1%; Pred. No. 1,6e-26;
Matches 111; Conservative 47; Mismatches 131; Indels 93; Gaps 14;

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:02:49 ; Search time 41 Seconds
(without alignments)
855.831 Million cell updates/sec

Title: US-09-831-426C-2
Perfect score: 2008
Sequence: 1 MDPVAVVAESVSSLTIAADF.....ESPNCVEDKMLSTFAVLTIG 365

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1992	99.2	363	2	DNA/RNA-binding pr
2	1814.5	90.4	338	2	transcription fact
3	1051.5	52.4	344	1	transcription fact
4	1047	52.1	335	2	transcription fact
5	1031	51.3	339	2	transcription fact
6	1022	50.9	339	2	transcription fact
7	556.5	27.7	365	2	5S RNA-binding pro
8	553.5	27.6	365	2	5S RNA-binding pro
9	496	24.7	374	2	zinc-finger protei
10	409	20.4	336	2	zinc-finger protei
11	406	20.2	693	2	zinc-finger protei
12	404	20.1	542	2	zinc-finger protei
13	398.5	19.8	626	2	zinc-finger protei
14	397.5	19.8	591	2	zinc-finger protei
15	394.5	19.6	803	2	zinc-finger protei
16	392.5	19.5	337	2	zinc-finger protei
17	391	19.5	651	2	zinc-finger protei
18	386.5	19.2	457	2	zinc-finger protei
19	386.5	19.2	615	2	zinc-finger protei
20	385	19.2	600	2	zinc-finger protei
21	382.5	19.0	280	2	zinc-finger protei
22	381.5	19.0	485	2	zinc-finger protei
23	380.5	18.9	595	2	zinc-finger protei
24	377.5	18.8	595	2	zinc-finger protei
25	375.5	18.7	496	2	zinc-finger protei
26	375	18.7	710	2	zinc-finger protei
27	374.5	18.7	515	2	zinc-finger protei
28	374	18.6	1350	2	zinc-finger protei
29	374	18.6	701	2	zinc-finger protei

30	371	18.5	728	2	probable transcrip
31	371	18.5	1191	2	zinc finger protei
32	368.5	18.4	654	2	transcription fact
33	368.5	18.4	673	2	transcription fact
34	368	18.3	435	2	zinc-finger protei
35	367.5	18.3	476	2	zinc-finger protei
36	367	18.3	675	2	zinc-finger protei
37	366.5	18.3	428	2	zinc-finger protei
38	365.5	18.2	399	2	zinc-finger protei
39	365	18.2	580	2	zinc-finger protei
40	364.5	18.2	589	2	zinc-finger protei
41	363.5	18.1	393	2	zinc-finger protei
42	363.5	18.1	469	2	zinc-finger protei
43	363	18.1	594	2	zinc-finger protei
44	362.5	18.1	553	2	zinc-finger protei
45	362.5	18.1	688	2	zinc-finger protei

ALIGNMENTS

RESULT 1	138937	DNA/RNA-binding protein - human (fragment)	C:Species: Homo sapiens (man)	C:Date: 29-May-1998	#sequence=1	revision=29-May-1998	#text=change 21-Jul-2000
C:Accession: 138937	R: Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Bliddison, W.E.; Becker, K.G.	Gene 159, 215-218, 1995	A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA	A:Reference number: 138937; MUID:95347600; PMID:7622052	A:Accession: 138937	A:Status: preliminary; translated from GB/EMBL/DBJ	A:Molecule type: mRNA
A:Residues: 1-363	A:Cross-references: EMBL:020272; NID:9644870; PIDN:AAV75623.1; PID:9644871	C:Suprafamily: transcription factor IIA	Query Match	99.2%	Score 1992;	DB 2;	Length 363;
Best Local Similarity	99.7%	Pred. No. 1.2e-143;	Matches 362;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	3	PPVAVVAESVSSLTIAADAFIAGESSAPTPPPALPFRFICSPFDCSANYSKAWKLDANHC	62	1	PPVAVVAESVSSLTIAADAFIAGESSAPTPPPALPFRFICSPFDCSANYSKAWKLDANHC	60	
Db	1	PPVAVVAESVSSLTIAADAFIAGESSAPTPPPALPFRFICSPFDCSANYSKAWKLDANHC	60	63	KATGERPVCDEYEGCGKAFIRDYHLSRHILTHTEGKPFVCAATGCDOKFNTKSNLKKHFE	122	
Qy	63	KATGERPVCDEYEGCGKAFIRDYHLSRHILTHTEGKPFVCAATGCDOKFNTKSNLKKHFE	122	61	KATGERPVCDEYEGCGKAFIRDYHLSRHILTHTEGKPFVCAATGCDOKFNTKSNLKKHFE	120	
Db	61	KATGERPVCDEYEGCGKAFIRDYHLSRHILTHTEGKPFVCAATGCDOKFNTKSNLKKHFE	120	123	RKHENQKQYICSEFEDECKTKFKKHQDLKHOCQHTNEPLFKCTQDECCGHNFPSPKLRH	182	
Qy	123	RKHENQKQYICSEFEDECKTKFKKHQDLKHOCQHTNEPLFKCTQDECCGHNFPSPKLRH	182	121	RKHENQKQYICSEFEDECKTKFKKHQDLKHOCQHTNEPLFKCTQDECCGHNFPSPKLRH	180	
Db	121	RKHENQKQYICSEFEDECKTKFKKHQDLKHOCQHTNEPLFKCTQDECCGHNFPSPKLRH	180	183	AKAHGIVYVOCGSGFPAKTWTTLKTHVRETHKEIICEVCRKTFKKDKVILKQMKHNAE	242	
Qy	183	AKAHGIVYVOCGSGFPAKTWTTLKTHVRETHKEIICEVCRKTFKKDKVILKQMKHNAE	242	181	AKAHGIVYVOCGSGFPAKTWTTLKTHVRETHKEIICEVCRKTFKKDKVILKQMKHNAE	240	
Db	181	AKAHGIVYVOCGSGFPAKTWTTLKTHVRETHKEIICEVCRKTFKKDKVILKQMKHNAE	240	241	RDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAAGCGKTFAMKQSLTRHVAVD	300	
Qy	241	RDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAAGCGKTFAMKQSLTRHVAVD	300	243	RDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAAGCGKTFAMKQSLTRHVAVD	302	
Db	243	RDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAAGCGKTFAMKQSLTRHVAVD	302	303	PDKKKKKLKVKRSRKRSLASHISGTYIPPKRKGGLSLCQNGESPNCEVDKMLSTVAVL	362	
Qy	303	PDKKKKKLKVKRSRKRSLASHISGTYIPPKRKGGLSLCQNGESPNCEVDKMLSTVAVL	362	301	PDKKKKKLKVKRSRKRSLASHISGTYIPPKRKGGLSLCQNGESPNCEVDKMLSTVAVL	360	
Db	301	PDKKKKKLKVKRSRKRSLASHISGTYIPPKRKGGLSLCQNGESPNCEVDKMLSTVAVL	360	Qy	363	TLG 365	
Qy	363	TLG 365		Db	361	TLG 363	
Db	361	TLG 363		RESULT 2	G01496		

transcription factor IITA - northern leopard frog
C:Species: Rana pipiens (northern Leopard frog)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C:Accession: JCI1441

R:Gaskins, C.J.; Smith, J.F.; Ogilvie, M.K.; Hanas, J.S.
Gene 120, 197-206, 1992

A>Title: Comparison of the sequence and structure of transcription factor IITA from Bufo terrestris and Rana pipiens.

A:Reference number: JC1441; MUID:93013035; PMID:138134

A:Accession: JC1441

A:Molecule type: mRNA

A:Residues: 1-335 <GAS>

C:Cross-references: EMBL:X58369; NID:g429060; PIDN:CAA11260.L; PID:g429061

C:Genetics:

A:Gene: TFIIA

C:Superfamily: transcription factor IITA

C:Keywords: DNA binding; duplication; nucleus; transcription factor; zinc finger

Query Match 52.1%; Score 1047; DB 2; Length 335;
Best Local Similarity 62.1%; Pred. No. 4,7e-72;
Matches 187; Conservative 32; Mismatches 76; Indels 6; Gaps 2;

Oy 34 PALPRFICSPDGCANYSKAMKLDAHLCKHTGEHPFYCDYEGCGKAFIRDYHLSRHILT 93
|| : ||| | ||| : || | ||| ||| ||| ||| : ||| : ||
Db 7 PAVVRRITCSFRADSCASATNKMKLDLAHLCKHTGEHPFCITYEGCGKGVTLEHLIRHSMT 66

Oy 94 HTGEKFVCYAATGGCDOKRNTFSNLKKHFERRKENOXOYLISFEDECKTFKKHOOLIKHQ 153
|||| | ||| | ||| : ||| : ||| : ||| | ||| | ||| |
Db 67 HTGEKPCKCADPDDSLSTTMTNLKHQRAHLPSLSLYECLFDCCGTFFKNQDLKHQ 126

Oy 154 COHNTEPLFKCTQSGCGGNAPSRLKRHAKAHEGYCOK--GCSFAVKTWELIKHVRE 211
||| : ||| ||| ||| : ||| : ||| : ||| | ||| : ||| : ||| : |||
Db 127 YIHTNOGFCKNHCECDGSFPSSPSRLKRHKRYNAAGYPQDCSSCFVGKTWETVEYKHAA 186

Oy 212 THKEKITLEDVCRKTPRKRDYLLKOHMKTTHAREVDVRCRGREGCGRTYTIVNIQSIIISFH 271
: || : ||| : || : ||| : ||| ||| : ||| : ||| ||| | ||| ||| |||
Db 187 SHSEPTTCDVCAKRFKKNTHTLKDHRKTHEVERVYKCPDGDRYYTKKFGLOSHGISFH 246

Oy 272 ESERPFVECHAGCGFTFMAMKSILTRHAVVHDPPDKKMKLKYSREKRSLSASHLSGIIP 331
:||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 EDNRFPACGHPCGCGTTFMKGOSLDHANATHPEEKKM-----KKRPKKSLSARLSGLGYPK 302

Oy 332 K 332
|
Db 303 K 303

RESULT 5
B34895
transcription factor IITA - Kenyan clawed frog
C:Species: Xenopus borealis (Kenyan clawed frog)
C:Date: 20-Jul-1990 #sequence_revision 21-Oct-1992 #text_change 26-Aug-1999
C:Accession: B34895; S09653
R:John, K.E.; Dairdy, M.K.; Crawford, E.T.; Brown, D.D.
Cell 61, 293-300, 1990

A>Title: A finger protein structurally similar to TFIIA that binds exclusively to 5S RNA

A:Reference number: A34895; MUID:90235278; PMID:2331751

A:Accession: B34895

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-339 <JOH>

A:Cross-references: GB:M32472; NID:g214150; PIDN:AAA49713.1; PID:g214151

R:Gaskins, C.J.; Hanas, J.S.
Nucleic Acids Res. 18, 2117-2123, 1990

A>Title: Sequence variation in transcription factor IITA.

A:Reference number: S09653; MUID:90245658; PMID:2110661

A:Accession: S09653

A:Molecule type: mRNA

A:Residues: 1-221,'C','223-234','E','236-291','R','293-309','S','311-313','A','315-318','G','320-333'
C:Cross-references: EMBL:X17695; NID:664474; PIDN:CAA35689.1; PID:g64475

C:Superfamily: transcription factor IITA

C:Keywords: DNA binding; nucleus; transcription factor; zinc finger

Query Match 51.3%; Score 1031; DB 2; Length 339;

[illegible]

Db 302 KSKKKKPSQTPAMSEDOOPDASKADPLVLENTLTL 338

RESULT 7

55 RNA-binding protein p43 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 20-Jul-1990 #sequence_revision 21-Oct-1992 #text_change 21-Aug-1998

C:Accession: A34895

R:John, K.E.; Darby, M.K.; Crawford, E.T.; Brown, D.D.

Cell 61, 293-300, 1990

A:Title: A finger protein structurally similar to TFIITa that binds exclusively to 5S rRNA

A:Reference number: A34895; MUID:90235278; PMID:2331751

A:Accession: A34895

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-365 <JON>

A:Superfamily: transcription factor IIIA

C:Keywords: DNA binding; zinc finger

Query Match 27.7%; Score 556.5; DB 2; Length 365;

Best Local Similarity 38.1%; Pred. No. 7.9e-35;

Matches 106; Conservative 48; Mismatches 121; Indels 3; Gaps 3;

42 CSPPDCSANYSKAMKLDANLCKHTGERPVCDEGCGKAFIRIDYHLSRHILTHNGEKPEV 101

17 CPAGACGKAFYREKGLQDMAGHSEQKPMKCGITKDCDDKVFARKRQILKHKVRLALKLS 76

102 CAATGCDGKFNKSNLKKHFERKHEKNOQKQYICSEFDDCKTKTKHQOLKIHCCQHTNEPL 161

77 CPTACCKMTFSTKSLSRKLYKH-GEAVPLKCFYPCGSRFRKRLRLSLVSNSEPL 135

162 FCTGEGCGKHNFAASPTKLRHAKHEGYVQ-KGCSFVAKTWTELLKHNRETHKEEILCE 220

136 SVCDVPGCGKSSVAKLVANHOKRHGYCSYEGCOTVSPYALQTHYK-HPLELQCA 194

221 VCRKTFKRKDYLOKMKTHAPERYDCRPRECGGRTYTTVFNLOSHILSFHEESRPVCE 280

195 ACKKPEKKSALRRHAKTHAKKPLDPCPRQDCDTESSVFNLTHVRYVHNLQTHRCR 254

281 HAGCGTFAMKOSLTRHNVHDPDKKMKLKVKRSREK 318

255 HSGCTKSFAMRESLLRHLVHDPERKIKLKLFVRPSK 292

RESULT 8

55 RNA-binding protein p43 - Kenyan clawed frog

C:Species: Xenopus borealis (Kenyan clawed frog)

C:Date: 20-Jul-1990 #sequence_revision 21-Oct-1992 #text_change 21-Aug-1998

C:Accession: A34895

R:John, K.E.; Darby, M.K.; Crawford, E.T.; Brown, D.D.

Cell 61, 293-300, 1990

A:Title: A finger protein structurally similar to TFIITa that binds exclusively to 5S rRNA

A:Reference number: A34895; MUID:90235278; PMID:2331751

A:Accession: A34895

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-365 <JON>

A:Superfamily: transcription factor IIIA

C:Keywords: DNA binding; zinc finger

Query Match 27.6%; Score 553.5; DB 2; Length 365;

Best Local Similarity 37.4%; Pred. No. 1.3e-34;

Matches 108; Conservative 48; Mismatches 130; Indels 3; Gaps 3;

31 PRPALPRRFICSPDCSANYSKAMKLDANLCKHTGERPVCDEGCGKAFIRIDYHLSRH 90

6 PRPSKSYQFRCPAGACGKAFYREKGLQDMAGHSEQKPMKCGITKDCDDKVFARKRQILKHKVRLALKLS 65

91 ILTHNGEKPEVCAATGCDGKFNKSNLKKHFERKHEKNOQKQYICSEFDDCKTKTKHQOLK 150

66 MKNHLTLKHSCTPACCKMTFSTKSLSRKLYKH-GEAVPLKCFYPCGSRFRKRLRLSLVSNSEPL 124

QY 151 IHCCQHTNEPLFCTQEGCGKHNFAASPTKLRHAKHEGYVQ-KGCSFVAKTWTELLKHN 209

125 IHVSEHSNEPLSVCDVPGCGKSSVAKLVANHOKRHGYCSYEGCOTVSPYALQTHYK-HPLELQCA 184

QY 210 RETHKEEILCEVCKRTYFRKDYLOKMKTHAPERYDCRPRECGGRTYTTVFNLOSHILS 269

165 KK-HPLELQCAACKKPFKKSALRRHAKTHAKKPLDPCPRQDCDTESSVFNLTHVRYVHNLQTHRCR 243

QY 270 FHEESRPVCEHAGCGTKFAMKOSLTRHNVHDPDKKMKLKVKRSREK 318

244 VHLCLQTHRCRPSNCTRSFAMRESLLRHLVHDPERKIKLKLFVRPSK 292

RESULT 9

737676

zinc-finger protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000

C:Accession: T37676

R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.

submitted to the EMBL Data Library, October 1999

A:Reference number: 221736

A:Accession: T37676

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-374 <MCD>

A:Cross-references: EMBL:AL13675; PDB:CAB59689.1; GSPDB:GN00066; SPDB:SPAC144.09C

A:Experimental source: strain 972h; cosmid c144

C:Genetics:

A:Gene: SPDB:SPAC144.09C

A:Map position: 1

A:Intons: 50/3

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 24.7%; Score 496; DB 2; Length 374;

Best Local Similarity 33.2%; Pred. No. 3.1e-30;

Matches 109; Conservative 48; Mismatches 123; Indels 48; Gaps 12;

40 FICSPDCSANYSKAMKLDANLCKHTGERPVCDEGCGKAFIRIDYHLSRHILTHNGEK 99

23 FICPVEECCKKTSRBSLLQHLTHSNERPVCIDTYGCKAKYRKSHLKHRCHTNVAP 82

100 FVCAATGCDGKFNKSNLKKHFERKHEKNOQKQYICSEFDDCKTKFKHQOLKIH-CCQHTN 158

83 FSGHYDCDAQFYTOQHLEBRHIE--VNRKRPYACTWEGCDEGCSHQOLRSHISACHN 140

159 EPLFKCTGCGCGKHNFAASPTKLRHAKHEGYVQ-KGCSFVAKTWTELLKHNRETHKEEILCE 220

141 LRPYCTYDDCELRFAFKYKOKLQNHVRAHEKILISCPHESGVGHEGF--EKWSQLQNH 198

QY 210 RETHKEEILCEVCKRTYFRKDYLOKMKTHAPERYDCRPRECGGRTYTTVFNLOSH 266

199 REAHVVS--CSICGRQFKTAHLRHNVVLAHQTTLERKTYHCHMECKKSFPTSSALKKH 256

QY 267 ILSFHEESRPVCEHAGCGTKFAMKOSLTRH-----AVV 300

257 ISVIEGNNMAFHCD--SCGTFKGYKMLQRLHRLERTGCKRAHKRYINEGCIKHDGIEGVAI 314

QY 301 HDPDKKMKLKVKRSREKSLASHLSGY 328

315 HDQKEKELSNLVSDVAKK-IINEVYGH 341

RESULT 10

506578

finger protein (clone X1CGF57-1) - African clawed frog (fragment)

C:Species: Xenopus laevis (African clawed frog)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 31-Dec-1993

C:Accession: S06578

R:Heidefeld, W.; El-Baradi, T.; Mantel, H.; Pieler, T.; Koester, M.; Poeling, A.; Kno

J. Mol. Biol. 1989, 639-659, 1989

A:Title: Second-order repeats in Xenopus laevis finger proteins.

Tue, Feb 11 13:12:49 2003

us-09-831-426c-2.rpr

Page 7

```

0y 316 REKSLAHSLSGYLPKPKKQGLSLONGSP-----NCVEDKMLST 358
      : | : | : : | : | : | : | : |
Db 678 GAAFLSLSTLSTH-----KIHTGKPYKCEKCGAENRPSNLLHKH 723

```

Search completed: February 10, 2003, 17:48:19
Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 14:48:36 ; Search time 81 Seconds
(without alignments)
928.484 Million cell updates/sec

Title: US-09-831-426c-2
Perfect score: 2008
Sequence: 1 MDPAAVVAESVSLTADAF.....ESPNCVEDKMLSTVAVITLG 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1583.5	78.9	387	11	Q9CSH8
2	1582.5	78.8	400	11	Q9VHT7
3	1498	74.6	336	11	Q9VHT8
4	946.5	47.1	266	11	Q9CSW1
5	551.5	27.5	368	13	Q9W7D7
6	496	24.7	374	3	Q9UTL5
7	425	21.2	927	4	Q9UEG4
8	416.5	20.7	703	11	Q9RI61
9	414	20.6	411	13	P70043
10	406.5	20.2	403	5	Q9VXB4
11	406.5	20.2	559	11	Q91WM0
12	406.5	20.2	587	11	Q922X6
13	406.5	20.2	700	4	Q961R4
14	405	20.2	411	13	P70046
15	405	20.2	847	4	Q9H559
16	404	20.1	531	4	Q9H176

17	403	20.1	934	5	Q9VKE1	Q9VKF1 drosophila
18	402.5	20.0	756	5	Q91362	Q91362 drosophila
19	399	19.9	754	11	Q99J65	Q99J65 mus musculus
20	399	19.9	662	5	Q61360	Q61360 drosophila
21	398	19.8	891	5	Q61361	Q61361 drosophila
22	397.5	19.8	591	13	Q99046	Q99046 xenopus lae
23	397.5	19.8	626	4	Q8WUR9	Q8WUR9 homo sapien
24	397.5	19.8	720	6	Q9N003	Q9N003 macaca fasc
25	396	19.7	435	4	Q96BM3	Q96BM3 homo sapien
26	395	19.7	686	4	Q96RM4	Q96RM4 homo sapien
27	395	19.7	803	4	Q96T86	Q96T86 homo sapien
28	393	19.6	751	4	Q9NS43	Q9NS43 homo sapien
29	392.5	19.5	547	4	Q96OH6	Q96OH6 homo sapien
30	392.5	19.5	637	4	Q96IR2	Q96IR2 homo sapien
31	392.5	19.5	715	4	Q9UEX4	Q9UEX4 homo sapien
32	392	19.5	626	11	Q70230	Q70230 mus musculus
33	392	19.5	874	11	Q9RI63	Q9RI63 mus musculus
34	390.5	19.4	568	4	Q8TD23	Q8TD23 mus musculus
35	390.5	19.4	576	4	Q96ND2	Q96ND2 homo sapien
36	390.5	19.4	576	4	Q96ND2	Q96ND2 homo sapien
37	389	19.4	744	11	Q62788	Q62788 rattus norv
38	387	19.3	650	11	Q62886	Q62886 rattus norv
39	386.5	19.2	873	6	Q9XSR1	Q9XSR1 canis famli
40	386	19.2	578	4	Q96NH0	Q96NH0 homo sapien
41	385.5	19.2	569	4	Q96ND8	Q96ND8 homo sapien
42	385.5	19.2	595	4	Q8TF32	Q8TF32 homo sapien
43	385	19.2	600	13	Q91853	Q91853 xenopus lae
44	384.5	19.1	303	4	Q9BY31	Q9BY31 homo sapien
45	384.5	19.1	357	4	Q96JW6	Q96JW6 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9CSH8	PRELIMINARY:	PRT:	387 AA.
AC	Q9CSH8:			
DT	01-JUN-2001 (TREMBLrel. 17, last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)			
DE	2010015D03R1K protein (Fragment).			
GN	GTF3A OR 2010015D03R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;			
RA	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakura I.,			
RA	Salto T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Salto R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,			
RA	Kuenli P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombeerts P.,			
RA	Norione P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,			
RA	Hayashizaki Y.;			
RT	*Functional annotation of a full-length mouse CDNA collection.*;			
RL	Nature 409:685-690(2001).			
DR	EMBL: AK012797; BAB28476.1; -			
DR	HSSP: P03001; ITRF3.			

DR MGD:1913846; Gtf3a.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 9.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
DR DNA-binding; Zinc-finger.
KW NON_TER 1
FT 1
SQ SEQUENCE 387 AA; 43988 MW; A96F7E3F6288CA CRC64;

Query Match 78.9%; Score 1583.5; DB 11; Length 387;
Best Local Similarity 77.5%; Pred. No. 8,7e-139;
Matches 283; Conservative 32; Mismatches 47; Indels 3; Gaps 2;

OY 1 MDPRAVAESVSLTIAAFIAGESSAPTPRRPALPRPFICSPDSCASANYSKAKKLDAN 60
DB 24 LEPRVAEAVASSLTIAAFVAGCE--GPAPRRPALPRPFICSPDSCASANYSKAKKLDAN 81
OY 61 LCKHTGERPVCVDYEGCGKAFIRDYHLSRHLITHTGKRPVCAATGCDOKFNTKSNLKKH 120
DB 82 LCKHTGERPVCVDYEGCGKAFIRDYHLSRHLITHTGKRPVCAATGCDOKFNTKSNLKKH 141
OY 121 FERKHNQOQVYICSFEDCKTFKKHQOLKIHOCQHTNEPLFKCTQEGCGKHFASPSRLK 180
DB 142 IERKGNPOQOYVCSYEGCGKAFKKHQOLRTHOCQHTNEPLFKCTQEGCGKHFASPSRLK 201
OY 181 RHAKHNEGVCQKGSFPAKTTELLKHVRETHKEEILCEVCRTFKRKDYLOKMKHTHA 240
DB 202 RHGXVHEGYLCQKGSFPAKTTELLKHVRETHKEEILCEVCRTFKRKDYLOKMKHTHA 261
OY 241 PERDYRCRPGCGRTYTTVFNLQSHILSFHEESRPVCEHAGCGKTRAMKOSLTRNAVY 300
DB 262 PERDYRCRPGCGRTYTTVFNLQSHILSFHEESRPVCEHAGCGKTRAMKOSLTRNAVY 321
OY 301 HDPDKKKMLKVKRSRERKSLASHLSGYTPRKROGQGLSLQ--NGESPNCVEDKMLSTV 359
DB 322 HDPDKKKMLKVKRPRERSLASRLSGYFPRKROEPDYSLPNASESSSPEADLPPPA 381
OY 360 AVLTG 364
DB 382 ALLTV 386

RESULT 2

OBVHT7 PRELIMINARY: PRT; 400 AA.
ID 08VHT7
AC 08VHT7
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Transcription factor IIAA (Fragment).
GN GTF3A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21673987; PubMed=11814676;
RA Hanes J.S., Hocker J.R., Cheng Y.-G., Lerner M.R., Brackett D.J.,
RA Lightfoot S.A., Hanes R.J., Madhusudan K.T., Moreland R.J.;
RT "CDNA cloning, DNA binding, and evolution of mammalian transcription
factor IIAA."
RT Gene 282:43-52(2002).
RL EMBL: AF391798; AAL69686.1; -.
DR MGD: MGI:1913846; Gtf3a.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 9.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
FT NON_TER 1
SQ SEQUENCE 400 AA; 45417 MW; EDFFA8BD9D26EAFD CRC64;

Query Match 78.8%; Score 1582.5; DB 11; Length 400;
Best Local Similarity 77.5%; Pred. No. 1.1e-138;
Matches 283; Conservative 32; Mismatches 47; Indels 3; Gaps 2;

OY 1 MDPRAVAESVSLTIAAFIAGESSAPTPRRPALPRPFICSPDSCASANYSKAKKLDAN 60
DB 37 LEPRVAEAVASSLTIAAFVAGCE--GPAPRRPALPRPFICSPDSCASANYSKAKKLDAN 94
OY 61 LCKHTGERPVCVDYEGCGKAFIRDYHLSRHLITHTGKRPVCAATGCDOKFNTKSNLKKH 120
DB 95 LCKHTGERPVCVDYEGCGKAFIRDYHLSRHLITHTGKRPVCAATGCDOKFNTKSNLKKH 154
OY 121 FERKHNQOQVYICSFEDCKTFKKHQOLKIHOCQHTNEPLFKCTQEGCGKHFASPSRLK 180
DB 155 IERKGNPOQOYVCSYEGCGKAFKKHQOLRTHOCQHTNEPLFKCTQEGCGKHFASPSRLK 214
OY 181 RHAKHNEGVCQKGSFPAKTTELLKHVRETHKEEILCEVCRTFKRKDYLOKMKHTHA 240
DB 215 RHGXVHEGYLCQKGSFPAKTTELLKHVRETHKEEILCEVCRTFKRKDYLOKMKHTHA 274
OY 241 PERDYRCRPGCGRTYTTVFNLQSHILSFHEESRPVCEHAGCGKTRAMKOSLTRNAVY 300
DB 275 PERDYRCRPGCGRTYTTVFNLQSHILSFHEESRPVCEHAGCGKTRAMKOSLTRNAVY 334
OY 301 HDPDKKKMLKVKRSRERKSLASHLSGYTPRKROGQGLSLQ--NGESPNCVEDKMLSTV 359
DB 335 HDPDKKKMLKVKRPRERSLASRLSGYFPRKROEPDYSLPNASESSSPEADLPPPA 394
OY 360 AVLTG 364
DB 395 ALLTV 399

RESULT 3

OBVHT8 PRELIMINARY: PRT; 336 AA.
ID 08VHT8
AC 08VHT8
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Transcription factor IIAA (Fragment).
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DWLEY; TISSUE=BRAIN;
RX MEDLINE=21673987; PubMed=11814676;
RA Hanes J.S., Hocker J.R., Cheng Y.-G., Lerner M.R., Brackett D.J.,
RA Lightfoot S.A., Hanes R.J., Madhusudan K.T., Moreland R.J.;
RT "CDNA cloning, DNA binding, and evolution of mammalian transcription
factor IIAA."
RT Gene 282:43-52(2002).
RL EMBL: AF391798; AAL69685.1; -.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 9.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
FT NON_TER 1
SQ SEQUENCE 336 AA; 38514 MW; 7084F95B74D168BA CRC64;

Query Match 74.6%; Score 1498; DB 11; Length 336;
Best Local Similarity 80.3%; Pred. No. 6.4e-131;
Matches 265; Conservative 23; Mismatches 42; Indels 0; Gaps 0;

OY 35 ALPRPFCSPDSCASANYSKAKKLDANLCKHTGERPVCVDYEGCGKAFIRDYHLSRHLITHT 94
DB 6 ALPRPFCSPDSCASANYSKAKKLDANLCKHTGERPVCVDYEGCGKAFIRDYHLSRHLITHT 65
OY 95 TGEKRPVCAATGCDOKFNTKSNLKKHFERKHNQOQVYICSFEDCKTFKKHQOLKIHOC 154

Db	66	TGEKRPVCAIDGCGNQKSTKSNLKKHIERHNENQKQIYVCEBCKKAFFKKHQQLRTHQC	125
Qy	155	QHTNEPLFKLCIQEBCGKHFPASPSKLKNAKANEHYVCGKCSFYAKWTTLKLVRETHK	214
Db	126	QHTNEPLFRCHBECCGKHFPASPSRLKRGKHGHEGYLCQKSCSFYKWTTLKLVRETHK	185
Qy	215	EEILCEVCRKTFKKRDKYLKQHKMKNHAPERDVCRCPRGCGRTYTYVTNLOSHILSFHEES	274
Db	186	EEVCTVCGKFFKKRDHLKQHKMKNHAPERDVCRCPRGCGRTYTYVTNLOSHILSFHEEK	245
Qy	275	RPPVCEHAAGCGKTFKAMKOSLIRHNVVDDPKKKMLKLVKKSAREKRSIASLISGCIYPPKK	334
Db	246	RPPVCEHAAGCGKTFKAMKOSLIRHNVVDDPKKKMLKLVVRRERKRSIASLISGCIYPPKK	305
Qy	335	QGQGLSTLCQNGESPNCVCEKMLSTVAIVLT	364
Db	306	QEPDQSLPNSTESSSSPEATLALAAALITV	335

RESULT 4			
Q9CSM1			
ID	Q9CSM1	PRELIMINARY;	PRT; 266 AA.
AC	Q9CSM1;		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	26101110IRIK protein (Fragment).		
GN	GF3A OR 26101110IRIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukushima Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,		
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov H., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Gustincic N.J., Holt C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald N., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,		
RA	Wyshka-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohutsuki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
DR	EMBL: AK011845; BAB27876.1; -		
DR	HSSP; P03001; 1PF3		
DR	MCD; MGI:1913846; Gtf3a.		
DR	InterPro; IPR000822; ZnF_C2H2.		
DR	SMART; SM00355; ZnF_C2H2; 6.		
DR	SMART; SM00355; ZnF_C2H2; 6.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.		
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.		
KN	DNA-binding; zinc-finger.		
FT	NON_TER		
FT	NON_TER		
SQ	SEQUENCE 266 AA; AF2C4738C5B35518 CRC64;		

Query Match	47.1%;	Score 946.5;	DB 11;	Length 266;
Best Local Similarity	77.7%;	Pred. No. 7.5e-80;		
Matches 167;	Conservative 16;	Mismatches 23;	Indels 9;	Gaps 2

[illegible]

RESULT	5	
Q9W7D7		
ID	Q9W7D7	PRELIMINARY; PRT; 368 AA.
AC	Q9W7D7;	
DT	01-NOV-1999 (TrEMBLrel. 12, Created)	
DT	01-NOV-1999 (TrEMBLrel. 12, last sequence update)	
DT	01-MAR-2002 (TrEMBLrel. 20, last annotation update)	
DE	RNA binding protein 42Sp43.	
OS	Oryzias latipes (Medaka fish).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
CC	Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;	
CC	Bonelliformes; Adrianchthyidae; Oryziinae; Oryzias.	
OX	NCBI_TaxID=8090;	
RM	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ORANGE-RED;	
RX	MEDLINE=20070874; PubMed=10602271;	
RA	Kanamori A;	
RT	"Systematic identification of genes expressed during early oogenesis in medaka."	
RL	Mol. Reprod. Dev. 55:31-36(2000).	
CC	-1 - COFACTOR BINDS A 2PE-2S CLUSTER (BY SIMILARITY).	
CC	EMBL: AF128814; AAD38911.2; -	
DR	HSSP: P03001; 1TF3.	
DR	InterPro: IPR000564; 2FE2S_ferredoxin.	
DR	InterPro: IPR000842; ZnF_C2H2.	
DR	Pfam: PF00096; ZnF_C2H2; 9.	
DR	SMART: SM00355; ZnF_C2H2; 9.	
DR	PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.	
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.	
DR	PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 6.	
DR	DNA-binding; Iron-sulfur Met; Metal-binding; Zinc-finger.	
SO	SEQUENCE 368 AA; 42335 MW; A0A84CF04C3D144D CRC64;	

Query Match	27.5%;	Score 551.5;	DB 13;	Length 368;
Best Local Similarity	37.9%;	Pred. No. 4.9e-43;		
Matches 11;	Conservative 36;	Mismatches 139;	Indels 7;	Gaps 4;

Qy	40	PISEPPDCSANYSKAMLAHLILKHNESPEYCDYGKAFIDYHLSHILTHNGEX	99
Db	16	FTCHAGGAGCAPREMKLAHEVTHNGERCAPSPAGGSLFKTSHLKHVYLQHGVSX	75
Qy	100	FVCAATGCDCKENTKSMLKHNPERKHENOQOYICSPEDCKKTKRKHQOLKHOCOTHNE	15
Db	76	FOCFAPCAKASFLIDAOPLKKNHONASGHN - FKKNOPKCSLSFKRRRLKLNKEHNVH	13
Qy	160	PLKCTOEGGKHNASPSKLRHAKHNEVYC - OKGCSFVAKTTELLKHYREHKEEL	21
Db	134	PNFCSNIRCTAFPSHIAKRAHEKKNHAYSCPHNDCCOVVHTWSKLOHRLAK - HVSFT	19
Qy	219	CEVCRKTFKRRDYLOKHMKHAPERDVCSPREGCGPRTYVJNOSHLSFHNESRPV	27
Db	193	CGVEKUYDKAGALRRKRLHNSHKVYLCLSPRANCOAFPTTFPNDHNIRKVNHLQILKYR	25
Qy	279	CEHAGGCKTEAMKOSILTRHAAVNDPDKKMKLKVKRSREKRSLSHLSGYIPR	331

Db 253 CFPDCCPRTFVNRSMHRMHHDDPSAITVK---KFORPKSMOKRLMGANOP 302

RESULT 6

09UTL5 PRELIMINARY: PRT: 374 AA.

AC 09UTL5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Zinc-finger protein (Transcription factor I11A).
GN SPAC144.09c OR SFC2.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Davis P.,
RA Churcher C.M.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RM [2]
RA SEQUENCE FROM N.A.
RA Schulman D.B., Setzer D.R.;
RT "Identification and Characterization of Transcription Factor I11A from
RT Schizosaccharomyces pombe";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AL152675; CAB59689.1; -;
DR EMBL: AY091590; AAM0046.1; -;
DR HSSP: P07248; IARD.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf_C2H2; 10.
DR PRODOM: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 374 AA: 43851 MW: 74690701FF08FF6 CRC64:

Query Match 24.7%: Score 496; DB 3; Length 374;
Best Local Similarity 33.2%: Pred. No. 7e-38;
Matches 109; Conservative 48; Mismatches 123; Indels 48; Gaps 12;

QY 40 FICSPFDCSANSKAMKIDAHNCKHTGRRFYCDYEGCGKAFIRDYHLSRHLITHTGKRP 99
DB 23 FICSPYEBCKKTSRPSLSLQHLRTSHNERPFYCDYEGCGKAFYRKSHLTKIRKCHTNVPR 82
QY 100 FVCAATGCDQFNFTKSNLKKHFERKHNENQOYICSFEDCKKTFKKHQDLKIH-OCQHTN 158
DB 83 FSGCHVDGCGAAGYTOOHLEHNE--VHKRPYACTWECDCDFSKHQDLRSHISACHTN 140
QY 159 EFLFKCTQEGCGKHPASPSKLKRNA-KANE---GYVC-QKGC---SVAKWTETELKRV 209
DB 141 LRPYCTYDCLRLRATKQKLOHNHVAHEKIIISCPHESVCGHEGF--EKMSOLDNHI 198
QY 210 RETHNEELICEVCRKTFKKKDYLKQMKTH---APERDVCRRPGCGGTYTIVNLOSH 266
DB 199 KRAHNPSS--CSICGHPFKTAALRHNVVLHQTTLERKTYHCPRMGCKKSPTRSSALKRN 256
QY 267 ILSPHESHRFVCEHAGCGKTFAMKOSLTRH-----AVV 300
DB 257 ISVHIGNNAFHCDD--SCGTRFGYKHMLORHLERGTCCKAHKRPYINEGIRKHDIEGVAI 314
QY 301 HDPDKKKMKLVKKSREKRSLSHLSGY 328
DB 315 HDQKEKELSSNLVSDYAKK-IINEVTGH 341

RESULT 7

09UEG4

ID 09UEG4 PRELIMINARY: PRT: 927 AA.

AC 09UEG4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE KIAA0326 protein (Fragment).
GN KIAA0326.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RX [1]
RN SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RC MEDLINE-97349984; PubMed-9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.;
RL DNA Res. 4:141-150(1997).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AB002324; BAA20784.1; -;
DR HSSP: P08047; ISF2.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf_C2H2; 19.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2; 19.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 19.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 927 AA: 102756 MW: 176078A9CA407F8C0 CRC64:

Query Match 21.2%: Score 425; DB 4; Length 927;
Best Local Similarity 32.0%: Pred. No. 7.6e-31;
Matches 116; Conservative 42; Mismatches 115; Indels 89; Gaps 19;

QY 3 PRAVAESVSLTIADAFIAGESSAPTPRPRA-LPRRPGCFPPDCSANSKAMKIDAHN 61
DB 183 PRVVPANESLSRELVOG-----RPAGAEKPYTC--ECGKSFSSMSKLRHQ 227
QY 62 CNHTGERPVCYDEGCGKAFIRDYHLSRHLITHTGKRPVC----- 102
DB 228 RHHTGERPNTC--SECGKSFQSSHLVQHQRHTHTGKRPKPCDCKGCFSSWSNLVQHQT 285
QY 103 -----AATGCDQFNFTKSNLKKHFERKHNENQOYICSFEDCKKTFKKHQDLKIHQCO 155
DB 286 HTGKERPYKCTEKEKAFSTQNLTKH-QRSHNG-EKPYKCG--ECRRARVRSDDLQHOAT 341
QY 156 HTNEPLFKCTQEGCGKHPASPSKLKRNAKANEG---YVCQK-GCSFV-----AKTW 203
DB 342 HTGKERPYKCP--CGKRGQGNHNLKQKIHNGEKRPYCTEGCKSFQSSSELTOQRHTT 399
QY 204 -----ELKNVRETHKEE--ILCEVCRKTFKKKDYLKQMKTHABER 243
DB 400 GEKPYRELECGKSGFHSSTLTKHOR-THLRDPFCRPOCGKFTLSATILRHQRHTGER 458
QY 244 DVCRRPRECCGCTTYTIVNLOSHILSFHESSRPVCEHAGCGKTFAMKOSLTRHNVNDR 303
DB 459 PY-KCPK--CGKSFVSSNLNH-QRIHNGRPYIC--ADCGKSFMSSTLIRHORHTG 512
QY 304 DK 305
DB 513 EK 514

RESULT 8

09RI61

09RI61 PRELIMINARY: PRT: 703 AA.

AC 09RI61;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Zinc finger protein ZFP235.
 GN ZFP235.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hauser L., York M., Shannon M., Stubbs L.;
 RT "Differential expansion of homologous zinc-finger gene families in
 human chromosome 19q13.2 and mouse chromosome 7.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).
 DR EMBL: AF167321; AAD45930.1; -.
 DR HSSP: P25490; 12NM.
 DR MGD: MGI:1929117; Zfp235.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf-C2H2; 15.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; ZnF_C2H2; 12.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; ZnF_C2H2; 15.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 15.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
 DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 SQ SEQUENCE 703 AA; 79398 MW; 1A1DA000D10D7386 CRC64;

Query Match 20.7%; Score 416.5; DB 11; Length 703;
 Best local similarity 33.6%; Pred. No. 3.4e-30;
 Matches 115; Conservative 42; Mismatches 142; Indels 43; Gaps 18;

QY 45 PDCSANYSKAMKIDALHCKNHTGERPVCYEGCGKAFIRDYHLSRHILHTHGERPVC 104
 DB 315 PEGCKSFNOSHLVHLPIHTGKPYCCD--NCGKFSKSTDLINHCRAVHTGKPYKCEV 372
 QY 105 TCGDOKFNTKSKHFERKHNOKOYICSPFDCKKTFKHQOLITHCQHNPEPLFC 164
 DB 373 --CGKGFORSILOAH-ERHTG-ERPKCG--DCKRFSNSSLNTHQOVNHEEKPYKC 426
 QY 165 TOEGCGHNPSPSKLRHAKAHNG---YVQKGSFVAKTWELTKHVR-ETKKEITICE 220
 DB 427 DE--CGKRFSLNLSHQVHTGKPYCEE--CGKGSASSFSQSHQVHTGKPYCCS 483
 QY 221 VCKTKRKRDYLRQNMKTAPRDVCRCPREGCGRTYTVFNLSHLSFHEESRPVCE 280
 DB 484 VCGKFSQSSEYFQAHORVHTGKPY-RC--DVCGRFNMSLNHNH-QRVHTGKPYKCE 539
 QY 281 HAGCGKTFPMKOSLTHNAVYHDPDKKKMLKYKKSREKSLASHL-----SGYIPK- 332
 DB 540 E--CGKGSASMLQAHOSHTGEC--PRKCNACQSRQASHLQAHQVHTGKPYKC 594
 QY 333 -----RKQGGSLCQNGESP-NCVE-DKMLSTVALVT 363
 DB 595 DTGKAFSQRSNQLQVHQIHTGKPKCEGCKEFSMSAGLT 636

RESULT 9
 P70043
 ID P70043 PRELIMINARY; PRT; 411 AA.
 AC P70043;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Zinc finger protein XFDL 141.
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 CC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Bourgignon C., Bellefroid E., Bouwmeester T., Rausch O., Pieler T.;
 RT "XFDL 141.";
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).
 DR EMBL: U65897; AAB07010.1; -.
 DR HSSP: P25490; 12NM.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF00096; zf-C2H2; 12.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; ZnF_C2H2; 3.
 DR SMART: SM00355; ZnF_C2H2; 12.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
 DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 SQ SEQUENCE 411 AA; 46615 MW; 0F3984747406A23D CRC64;

Query Match 20.6%; Score 414; DB 13; Length 411;
 Best local similarity 36.6%; Pred. No. 3.2e-30;
 Matches 107; Conservative 28; Mismatches 107; Indels 50; Gaps 15;

QY 41 ICSEFDCSANYSKAMKIDALHCKNHTGERPVCYEGCGKAFIRDYHLSRHILHTGK 100
 DB 125 VC--PHGKRSYRDKTKLTVLRIHTGETPVC--PECGKGFSDASPLKSHLSHTGK 180
 QY 101 VCAATGCDOKFNTKSKNLKHFERKHNOKOYIC-----SFED----- 138
 DB 181 VC--TECGKSFRRKCELSILO--VHTGVKQFVCTECGKSFRRSELNHTGKTPP 236
 QY 139 ---CKRTFKHNOOLKTHOCQHNPEPLFCYTOEGCGKHPASPRLKRRHAKAHNG---YVC- 191
 DB 237 CTCGGSFPAKKRLKRNQHMHTGKPYECHE--CGKQFLKSLKLNHLSHTGKPYC 294
 QY 192 QKCSFVAKTWELTKHVRHTKEE--ILCEVCKTKRKRDYLRQNMKTAPRDVCRCP 249
 DB 295 ECGEOLF--TWKLLNQHLSHTGERPVCSECGKSFRTKASIALCHHTGKPY-CP 350
 QY 250 REGCGRTYTVFNLSHLSFHEESRPVCEHAGCGKTFPMKOSLTHNAVYH 301
 DB 351 E--CGKGFDPSTLSNHSVR-HTGKPYFC--AECCYTYDSSTLNHNLKLH 397

RESULT 10
 O9YXB4
 ID O9YXB4 PRELIMINARY; PRT; 403 AA.
 AC O9YXB4;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
 DE CG9609 protein (LD35155p).
 GN CG9609.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 DE Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Ceinalker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriil J.F., Adguyani A., An H.-J., Andrews-Plankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harlis M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasmo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Meruliov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relneer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Fartin D., Frise E., George R.,
RA Gonzalez M., Guertin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclebo J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF003503; AAF48663.1; -;
DR EMBL: AY051849; AAK93273.1; -;
DR HSP: P03001; 1Tf6.
DR P1yase; PBN0030787; CG9609.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf_C2H2; 9.
DR SMART: SM00355; Znf_C2H2; 8.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 403 AA; 46614 MW; 4AD11f430E96662 CRC64;

Query Match 20.2%; Score 406.5; DB 5; Length 403;
Best Local Similarity 30.3%; Pred. No. 1.5e-29;
Matches 105; Conservative 55; Mismatches 144; Indels 43; Gaps 14;

QY 39 FICSPDSCANSYKAMKLDANILCKHTGERPFVCGEGCGKAFIRDYHLSRNI-LTR--- 94
DB 34 KACSMKPCKEATFTKRLDQIDRHEHYHNTGILKNACTEGCDKTYSTYHLKRIHRSIHNER 93
QY 95 --TGERPFVCAATGCDQKFTNKSNLKHKFERKHENOQOYICSFEDCKTKFKKHOOLKIH 152
DB 94 ESAAKTKVACALEECSKMFISVSNMTRHMRTHES-PKYVPCG--QCSAKFSQKLKLRH 150
QY 153 QC-QHTNEPLFKTOEGCGKHFASPSKLRHAKAHGVCQKSCSVATWTELLKHVIE 211
DB 151 EIREHTELEVPYCSK--CSRGPFQOQOCSHSPSKLYCS-PCSLQDFDKWTLTYTHCRD 207
QY 212 T-----HKEEILCEVCRFTFKRDYLKQHMKT--HAPERDVC---DCPRPGCGRTY 258
DB 208 SLHGKRRHK-----CDKCDSAFDKPSFLKRNHLEKKAQOTDECATSPFCNEBGSSTVS 263
QY 259 TVFNLSHLSLFSHESRPVCEHAGCGKTFAMKOSLTRHAV-VHDPKMKMLKLVKKSRE 317
DB 264 YLRNLQHMLTAH-SGRRECGALDGCRCFSSAQNLARHLNRHKGATKKEIAKAKKKK 322
QY 318 KSLSLHLSGTYLPKRRKCGCGSLCONGESPCNVEDKMLSTVAVLTL 364
* 3

DB 323 SKTGECKTKSTSRKRRRDAGRS-----KHSRLSLACLQL 358

RESULT 11

Q91WMO PRELIMINARY; PRT; 559 AA.
AC Q91WMO;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 64.2 kda protein.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN NCBI_TaxID=10090;
RP [1]
RC SEQUENCE FROM N.A.
RA TISSUE-EYE, AND RETINA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC014712; AAH14712.1; -;
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF01352; KRAB; 1.
DR P1am; PF00096; zf_C2H2; 14.
DR ProDom: PD000003; Znf_C2H2; 6.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_14.
KW DNA-binding; Hypothetical protein; Zinc-finger.
SQ SEQUENCE 559 AA; 64216 MW; C08F544C0E2A429C CRC64;

Query Match 20.2%; Score 406.5; DB 11; Length 559;
Best Local Similarity 36.2%; Pred. No. 2.2e-29;
Matches 98; Conservative 42; Mismatches 106; Indels 25; Gaps 14;

QY 40 FICSPDSCANSYKAMKLDANILCKHTGERPFVCGEGCGKAFIRDYHLSRNI-LTHTGERP 99
DB 279 YVCK-HCGKATYTYTLRAHNSHSGEKRYVCKH--CGKATSYSTLRAHNSHSGEKP 334
QY 100 FVCAATGCCQKAFNTKSNLKHFKERKHENOQOYICGFEDCKTKFKKHOOLKIHOCQHTNE 159
DB 335 YVC-HKSCKAFQSSYLRIH-KRTHTG-EKPYIC--KQCKAFARSSHLQIKRSHTEP 388
QY 160 PLFKCTOECCGKHFASPSKLRHAKAHG---YVC-QKGSFVAKTWTELLKHVRETHKE 215
DB 389 KRYVCKQ--CGKAFQSSYLHINQSHTEGKRYVCKQCGKAFTRSSHLDIHK-ITHTGER 445
QY 216 EILCEVCRFTFKRKDYLKQHMKTAPERD-VCRCPREGCGRTYTVFNLSHLSFHEES 274
DB 446 PYSCRKLGKAFTHSNYLIQIKRINHTGEKRYVCK---ECGKAFARSTSLIH-EGTHSGE 500
QY 275 RRVVCHAGCGKTFAMKOSLTRHAVVHDPK 305
DB 501 KRYVCKQ--CGKAFYTLSSLRDHDVHSEK 529
RESULT 12
Q922X6
ID Q922X6 PRELIMINARY; PRT; 587 AA.
AC Q922X6;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Zinc finger protein sll-6.
GN sll-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP [1]
RC SEQUENCE FROM N.A.
RA STRAIN-C57B6; TISSUE-SPLEEN;

RA O-Wang J.;
RT "zinc finger protein.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AB020542; BAA34724.1; -.
DR HSSP: P08047; ISP2.
DR InterPro: IPR001909; KRA6.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF01352; KRA6; 1.
DR Pfam: PF00096; zf-C2H2; 15.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; znf_C2H2; 6.
DR SMART: SM00355; znf_C2H2; 15.
DR SMART: SM00355; znf_C2H2; 15.
DR PROSITE: PS00805; KRA6; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 15.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 16.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc; Zinc-finger.
SQ SEQUENCE 587 AA; 67450 MW; 94B3416511BE919 CRC64;

Query Match 20.2%; Score 406.5; DB 11; Length 587;
Best Local Similarity 36.2%; Pred. No. 2.4e-29;
Matches 98; Conservative 42; Mismatches 106; Indels 25; Gaps 14;

OY 40 FICSPDCSANYSKAMKLDALHCKHTGERPFVCDYEGCGKAFTRDYHLSRHILTHGKRP 99
DB 307 YVC--HCGKAYTYTNTLRAHSHSGEKPYVCKH--CGKAYTSYSLRAHSHSGEKRP 362
OY 100 FVCAAGCGDCKNTSKLKHFERKHNQOQYICSEEDCKTKPKKHQOLKHOCQHTNE 159
DB 363 YVC--KHCGKAFQSSYSLRTH--KRTHTG--EKPYIC--KOCGKAFARSSHQIHKRSHTGE 416
OY 160 PLKRCQEGCGKHFASPSKLRHAKAHG---YVC--QKCGSFAKMTTELLKHYREHKE 215
DB 417 KPYVCKO--CGKAFQSSYSLRTH--KRTHTG--EKPYIC--KOCGKAFARSSHQIHKRSHTGE 473
OY 216 EILCEVCRKTFKRDYLLQHMKTTHAPERD--VCRCPREGCGRTYTYFNLSHLSFHEES 274
DB 474 PYSGKLCGKAFHTSNYLIQIHKRIHTEGKPYVCK---ECGKAFARSSLSLTH--EGTHSGE 528
OY 275 RPYVCEIAGCGKTFAMQSLTRHVVADPDK 305
DB 529 KPYVCKO--CGKAFLLSSLRHVDVHSEK 557
RESULT 13
O961R4
ID O961R4 PRELIMINARY; PRT; 700 AA.
AC O961R4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to zinc finger protein 224.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EYE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC007303; AA07303.1; -.
DR InterPro: IPR001909; KRA6.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF01352; KRA6; 1.
DR Pfam: PF00096; zf-C2H2; 19.
DR PRODOM: PD000003; znf_C2H2; 5.
DR PROSITE: PS00805; KRA6; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_18.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 19.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 700 AA; 80559 MW; 79F266EA18821193 CRC64;

Query Match 20.2%; Score 406.5; DB 4; Length 700;
Best Local Similarity 30.4%; Pred. No. 2.9e-29;
Matches 120; Conservative 45; Mismatches 137; Indels 93; Gaps 19;

OY 46 DCSANYSKAMKLDALHCKHTGERPFVCDYEGCGKAFTRDYHLSRHILTHGKRPVCAAT 105
DB 236 ECGKGFSSRSSTLVHCKHLSHSGEKPYNC--ECGKAFIASHLQEHQRIHTEGKPFCDT- 292
OY 106 GCDQKNTKSNLKHFERKHNQOQYICSEEDCKTKPKKHQOLKHOCQHTNEPLFKCT 165
DB 293 -CGKAFRRRSALNNHC--VYHTEGKPYKC--EDCGKCFQSSNLRTHQVHTEGKPYKC- 346
OY 166 OEGCGKHFASPSKLRHAKAHG---YVCO--KG-----CGSPV 198
DB 347 -ECCGKCFIOPSOFAHRRHHTGKRPYVCKGKGYIYSSFOAHGCVHTGKRPYCNCG 405
OY 199 AKTWTELLKH---VETNKEELICEVCRKTFKRDYLLQHMKTTHAPERDVCRCRREGCG 254
DB 406 GKSFPMKIHQVHLVYHTGKPYKCEYCGKAFQSSYSLKTHLKAHSVQKPF--KC--ECG 462
OY 255 RRTTYTFFNLQSHLSFHEESRPVCEHAG-----CGKTF 288
DB 463 QGFNQSSRLQIHLQ--IHTGKRPYKCEGCGFSRRADLKINCRHHTGKRPYCNCEGKVF 521
OY 289 AMQSLTRHVVADPDKKKKTLVKKSREKRSLSHL-----SGYIPPKKO--GQGLS 340
DB 522 SQASHLTHQVHSGEK---PFKCEGCGKFSRSALQHMKVHTEGKPYKCEGCGKGR 578
OY 341 LCON-----GESP-NCVE--DKMLSFVAVLTL 364
DB 579 WSLNLDMQHVHTGKRPYTCGECGKHFQSSQSLQ 613

RESULT 14
P70046
ID P70046 PRELIMINARY; PRT; 411 AA.
AC P70046;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE XFDL141.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OX Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97321550; PubMed=9178260;
RA Bellefroid E., Bourguignon C., Boumeester T., Rausch O., Blumberg B.,
RA Peller T.;
RT "Transcription regulation and alternative splicing of an early zygotic
RT gene encoding two structurally distinct zinc finger proteins in
RT Xenopus laevis.";
RL Mech. Dev. 63:99-108(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: U67077; AAB72012.1; -.
DR HSSP: P25490; IUBD.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 12.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; znf_C2H2; 3.
DR SMART: SM00355; znf_C2H2; 12.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 12.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 411 AA; 46530 MW; E7309FEBFEB9704 CRC64;

Query Match 20.2%; Score 405; DB 13; Length 411;
Best Local Similarity 35.6%; Pred. No. 2.2e-29;
Matches 104; Conservative 31; Mismatches 107; Indels 50; Gaps 15;

```
QY 41 ICSFPDCSANYSKAMKLDANLCKHTGERPVCDEGCGKAFIRDYHLSRHILTHTEGKRP 100
DB 125 VC--PHCGSKYRDKTKLTVALRIHTGETFPVC--PECGKGFRRDGSFLKSHLSHTGKRP 180
QY 101 VCAATGCDOKFNTKSNLKKHFEKKHENOQKQYIC-----SFED----- 138
DB 181 VC--TECGKSFRRKCELSHLDQ--VHTGVKQFVCTEGCKSFRDRSELNTHINVTGKTFP 236
QY 139 ---CKRTFKKHQOLKIHCQNTNEPLFKCTOEGCGKHFPASPKLKRHAKANEH---YVC- 191
DB 237 CTGCGSKSFAKKMLKRMQMTHTGKRPHECTE--CGKQFLEKSLERHILHSHTGKRPSCF 294
QY 192 QKCSFVAKTWTLELKHVRETHKEE--ILCEVGRKTFKRKYDKONHKTTHAPERDYCRCP 249
DB 295 ECGEQF---TWKHLQNHQSHTEGRRPVCSECGSKSYKTKASLALCHHITGKRPY-CP 350
QY 250 REGCGRTYTVFNLOSHLSFHESRRPVCENHGGCTPFAMKOSLTRHAYVH 301
DB 351 E--CGKGFDRPTSLNSHVR--HNGEKPFAC--AECGKTYTDSSTLKNHLKLH 397
```

RESULT 15

```
Q9H559 PRELIMINARY: PRT: 847 AA.
ID Q9H559;
AC Q9H559;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, last annotation update)
DE BA526D8.4 (Novel KRAB box containing C2H2 type zinc finger protein)
DE (Fragment).
GN BA526D8.4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AL136981; CAC12728.1; -.
DR HSSP: P07248; 1PAA.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; zf-C2H2; 17.
DR PRINTS: PR00048; ZINCFINGER.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; ZnF_C2H2; 17.
DR PROSITE: PS08085; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 18.
DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
FT NON_TER 1
FT SEQUENCE 847 AA; 97660 MW; A86DCA4CDD7541D CRC64;
```

Query Match 20.2%; Score 405; DB 4; Length 847;

Best local similarity 32.9%; Pred No. 4, 9e-29;

Matches 105; Conservative 42; Mismatches 118; Indels 54; Gaps 15;

```
QY 42 CSEPPDCSANYSKAMKLDANLCKHTGERPVCDEGCGKAFIRDYHLSRHILTHTEGKRPV 101
DB 437 CS--DCGKSFIRKKSQHLVHQRIHTGENPFC--SECGKVFTHKTNLTHQKHTGKRPYI 492
QY 102 CAATGCDOKFNTKSNLKKHFEKKHENOQKQYICSFEDCKTFKKHQQOLKIHCQNTNEPL 161
DB 493 C--TVCGKAFTRDSNLKIH--QKIHGKRPYKS--DCGKSFTWKSRLRIHQCHTGERH 546
QY 162 FKCTOEGCGKHFPASPKLKRHAKANEH--YVCQK--GCSPFAKTWTLELKHVR--ETHKEE 216
DB 547 YECSE--CGKAFIQKSTLSHQRHNGEKRPYVTECGKAFPHK--SHFTHERIHTEGKRP 602
QY 217 ILCEVGRKTFKRKYDKONHKTTHAPERDYCRPREGCGRTYTVFNLOSHLSFHESRP 276
```

```
DB 603 YECISGKSFTKKSQHLVHQRIHTGKRPY-RCAE--CGKAFTRDSNMLFTIH-QKIHTEGKRP 658
QY 277 FVCEHAG-----CGTFAMKOSLTRHAYVHDPDKKKM-- 308
DB 659 YKSDCGKAFTRKSGLIHQOSHTEGRRHYECSECGKAFARKSTLIMHQRHTEGKRPYICN 718
QY 309 ---KLKVKKSREKRSLSH 324
DB 719 ECGKSFQKSHLNRHRIH 737
```

Search completed: February 10, 2003, 17:47:29

Job time : 89 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:37:40 ; Search time 290.342 Seconds
(without alignments)
9873.841 Million cell updates/sec

Title: US-09-831-426c-3

Perfect score: 1273

Sequence: 1 atgcgcagcagcgcgccgca.....cagtactacccttgctaa 1273

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

N.Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1273	100.0	1273	21	AAAI5405	DNA encoding a hum
2	1233.2	96.9	1381	24	ABR83650	Human CDNA differe
3	1231.6	96.7	1399	17	AAAT14038	Transcription fact
4	1228.6	96.5	1269	17	AAAT14037	Transcription fact
5	1211.4	95.2	1213	21	AAAI5407	Fragment of DNA en
6	1206	94.7	1349	22	AAH34586	Human colon cancer
7	495	38.9	690	24	ABO54923	Human ovarian anti
8	418.4	32.9	439	24	ABR44775	CDNA encoding colo
9	416.8	32.7	439	24	ABR45138	CDNA encoding colo

10	374	29.4	537	24	ABO59499	Human colon cancer
11	301.4	23.7	466	23	ABV49231	Human prostate exp
12	300.4	23.6	425	23	ABV19459	Human prostate exp
13	246	19.3	1401	22	AAH22129	Human eukaryotic 1
14	236	18.5	1889	23	AAH73534	DNA encoding novel
15	234	18.4	234	24	ABK45059	CDNA encoding colo
16	193	15.2	291	16	AAAT19568	Human gene signatu
17	177	13.9	272	24	AAI69173	Activated T-cell d
18	152.4	12.0	660	22	AAH23370	Nucleotide sequenc
19	152.4	12.0	947	22	AAH14428	DNA encoding zinc
20	152.4	12.0	947	22	AAH89338	Plasmid pFIIIAI1f
21	152.4	12.0	947	22	AAAC91046	Vector pFIIIA/zif-
22	152.4	12.0	995	22	AAH14427	DNA encoding zinc
23	152.4	12.0	995	22	AAH89337	Plasmid pFIIIAI1f
24	152.4	12.0	995	22	AAAC91045	Vector pFIIIA/zif-
25	101.6	8.0	183	24	ABN24598	Human ORF polynuc
26	97	7.6	919	22	AAH48496	Japanese medaka fe
27	91.2	7.2	3302	24	ABL69125	Kidney cancer rela
28	81	6.4	368	24	ABK46196	CDNA encoding colo
29	80.4	6.3	2400	22	ABK48579	Human breast cell
30	80.4	6.3	2400	22	ABK65494	Human foetal liver
31	80.4	6.3	2400	22	ABA33554	Probe #12020 for g
32	80.4	6.3	2400	22	AAK14917	Human brain expres
33	80.4	6.3	2400	22	AAK40649	Human bone marrow
34	80.4	6.3	2400	22	AAI21411	Probe #11344 for g
35	80.4	6.3	2400	22	AAI46696	Probe #15382 used
36	80.4	6.3	2400	22	AAI07105	Probe #7096 used t
37	80.4	6.3	2400	24	ABH14609	Human genome deriv
38	78.8	6.2	2981	22	AAK53131	Human polynucleoti
39	75.6	5.9	3135	22	AAK52147	Human polynucleoti
40	74.6	5.9	1662	22	AAH34049	Human colon cancer
41	74.6	5.9	4017	23	ABV24327	Human prostate exp
42	74.2	5.8	1830	24	ABA96948	Human Spt family t
43	73.2	5.8	1197	24	ABH96949	Human Spt family t
44	67	5.3	1675	23	AAH79321	Human Spt family t
45	64.6	5.1	292	22	AAH57392	DNA encoding novel

ALIGNMENTS

RESULT 1	AAAI5405	standard; DNA: 1273 BP.
ID	AAAI5405	
AC	AAAI5405;	
XX		
DT	04-SEP-2000	(first entry)
XX		
DE	DNA encoding a human transcription factor designated htfiitA.	
XX		
KW	Human; transcription factor; htfiitA; DNA-binding protein;	
KW	transcription; ribosomal RNA 5S gene; transcriptional control;	
KW	cancer; ss.	
XX		
OS	Homo sapiens.	
XX		
PH	key	Location/Qualifiers
FT	CDS	176..1273
FT		/tag- a
XX		/product- "Transcription factor"
XX		
PN	WO200028024-A1.	
XX		
PD	18-MAY-2000.	
XX		
PF	09-NOV-1999;	99WO-FR02738.
XX		
PR	10-NOV-1998;	98FR-0014146.
XX		
PA	(HMRI) HOECHST MARION ROUSSEL.	
XX		
PI	Bordon-Pallier F, Roher C;	

PI	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J:	
XX	WPI: 2002-435328/46.	
XX		
PT	Detecting granulocyte activation by detecting differential expression	
PT	of genes associated with granulocyte activation, which serves as	
PT	diagnostic markers that is useful for monitoring disease states and	
PT	drug toxicity	
XX		
PS	Claim 1: SEQ ID NO 221; 114pp; English.	
XX		
CC	The invention relates to detecting (M1) granulocyte (GC) activation	
CC	(GCA), by detecting the level of expression of gene(s) (Gs). Identified by	
CC	DNA chip analysis as given in the specification, and comparing	
CC	the expression level to an expression level in an unactivated	
CC	GC, where differential expression of Gs is indicative of GCA.	
CC	Also included are modulating (M2) GA by contacting GC with an agent	
CC	that alters the expression of at least one gene in Gs; (2) screening (M3)	
CC	for an agent capable of modulating GCA or an inflammation (especially	
CC	chronic) in a tissue, an allergic response in a subject, exposure of a	
CC	subject to a pathogen or sterile inflammatory disease using the	
CC	gene expression profile; (3) detecting (M4) an inflammation (especially	
CC	chronic) in a tissue, an allergic response in a subject, exposure of a	
CC	subject to a pathogen or sterile inflammatory disease, by detecting the	
CC	level of expression in a sample of the tissue of gene(s) from Gs, where	
CC	the level of expression of the gene is indicative of inflammation;	
CC	(4) treating (M5) an inflammation (especially chronic) or in a tissue,	
CC	an allergic response in a subject, exposure of a subject to a pathogen	
CC	or sterile inflammatory disease, by contacting a tissue having	
CC	inflammation with an agent that modulates the expression of gene(s)	
CC	from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for	
CC	modulating GA; M3 is useful for screening an agent capable of modulating	
CC	GCA preferably in an inflammation in a tissue; M4 is useful for	
CC	detecting an inflammation (especially chronic) in a tissue, an allergic	
CC	response in a subject, exposure of a subject to a pathogen or sterile	
CC	inflammatory disease (e.g. psoriasis, rheumatoid arthritis,	
CC	glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal	
CC	reflexion injury, ARDS, adult respiratory distress syndrome,	
CC	inflammation bowel disease, Crohn's disease, ulcerative colitis,	
CC	periodontal disease; also bacterial infection, viral infection,	
CC	parasitic infection, protozoal infection, fungal infection and M5 is	
CC	useful for treating one of the above conditions. The present	
CC	sequence represents a gene differentially expressed in granulocytes.	
CC	Note: The sequence data for this patent did not form part	
CC	of the printed specification, but was obtained in electronic	
CC	format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pcl_sequences.	
XX		
XX	Sequence 1381 BP: 386 A; 350 C; 354 G; 291 T; 0 other;	
XX		
Query Match	96.9%; Score 1233.2; DB 24; Length 1381;	
Best Local Similarity	99.5%; Pred. No. 9.6e-312;	
Matches 1268: Conservative	0; Mismatches 3; Indels 3; Gaps 3.	
Y	1 ATGCGCAGCAGCGGGCGCGCGCGGGGGGCGTGCCTGTGACCGCGCGGCTCCGGAAGT 60	
DB	20 ATGCGCAGAGCGGGCGCGCGCGGGGGCGTGCCTGTGACCGCGCGGCTCCGGAAGT 79	
Y	61 GTGCGCGGGTCCCGCGGAGGTTTCACAGGAGACCGTGGCGCGCGCGCGGTTCGGCG 120	
DB	80 GTGCGCGGGTCCCGCGGAGGTTTCACAGGAGACCGTGGCGCGCGCGCGGTTCGGCG 138	
Y	121 ACGTCTTCGCGACGCTGAGCGAGCGCGCGTGGCGCTTGGAGGAGCGCGGCGCCTGGA 180	
DB	139 ACGTCTTCGCGACGCTGAGCGAGCGCGCGTGGCGCTTGGAGGAGCGCGGCGCCTGGA 198	
Y	181 TCCGCGCGCGCGTGTGCGCGAGTGGGTGTCCTCTTGACATACGCGCGACGCTTGATTTC 240	
DB	199 TCCGCGCGCGCGTGTGCGCGAGTGGGTGTCCTTGACATACGCGCGACGCTTGATTTC 258	
Y	241 ACGCGCGAGAGACTACGCTTCGACACCGCGCGCGCGCGCGCTTCCAGAGGTTCACTTG 300	
DB	259 ACGCGCGAGAGACTACGCTTCGACACCGCGCGCGCGCGCGCTTCCAGAGGTTCACTTG 318	

QY	301	CTTCTTCCCTGACATGCAGGCGCCCAATTACAGCAAAAGCTGTGAGAGCTTGAACGGCCACTGTG	360
Db	319	CTTCTTCCCTGACATGCAGGCGCCCAATTACAGCAAAAGCTGTGAGAGCTTGAACGGCCACTGTG	378
QY	361	CAACGACACGGGGGAGAGACATTTTGTGTGACTATGAAAGGGTGTGGCAAGGCTTCAT	420
Db	379	CAACGACACGGGGGAGAGACATTTTGTGTGACTATGAAAGGGTGTGGCAAGGCTTCAT	438
QY	421	CAGGAGCTACCATCTGAGCGCCACATTCATCACACAGAGAAAACCGTTTGTG	480
Db	439	CAGGAGCTACCATCTGAGCGCCACATTCATCACACAGAGAAAACCGTTTGTG	498
QY	481	TGCAGCCACTGGCTGTGATCAAAAATTCAACACAAATCAAACTTGAGAAACATTTTGA	540
Db	499	TGCAGCCCAATGGCTGTGATCAAAAATTCAACACAAATCAAACTTGAGAAACATTTTGA	558
QY	541	ACGCAAAACATGAAAATGCAAAAAACAAATATATGACATTTTGAAGACTGTAAAGAGAC	600
Db	559	ACGCAAAACATGAAAATGCAAAAAACAAATATATGACATTTTGAAGACTGTAAAGAGAC	618
QY	601	CTTTTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCAGCATTTACCAATGAAACCTCTATT	660
Db	619	CTTTTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCAGCATTTACCAATGAAACCTCTATT	678
QY	661	CAATGTATCCAGCAAGAGATGTGGGAAACACTTTGATACCCACCAAGCTGAAAACGCA	720
Db	679	CAATGTATCCAGCAAGAGATGTGGGAAACACTTTGATACCCACCAAGCTGAAAACGCA	738
QY	721	TGCCAAGGCCACAGAGGGCTATGTATGTCAAAAAGGATTTCTTGTGGCAAAAACATG	780
Db	739	TGCCAAGGCCACAGAGGGCTATGTATGTCAAAAAGGATTTCTTGTGGCAAAAACATG	798
QY	781	GACGGAACTTCGAAACATGTGAGAGAAACCCATAAGAGAAATCTATGTGAAGTATG	840
Db	799	GACGGAACTTCGAAACATGTGAGAGAAACCCATAAGAGAAATCTATGTGAAGTATG	858
QY	841	CCGGAAAAACATTTAAAGCAAAATTTACCTTAAAGCAACATGAAAACCTCANTGCCCGCA	900
Db	859	CCGGAAAAACATTTAAAGCAAAATTTACCTTAAAGCAACATGAAAACCTCANTGCCCGCA	918
QY	901	AAGGATGTATGTGCTGCTCAAGAGAGAGGCTGTGAAAGAACTATACTAGTGTGTTAA	960
Db	919	AAGGATGTATGTGCTGCTCAAGAGAGAGGCTGTGAAAGAACTATACTAGTGTGTTAA	978
QY	961	TCTCCAAAAGCCATATCCTCTCTCTTCCATGAGAAAGCCGCTTTTGTGTGTAACATGC	1020
Db	979	TCTCCAAAAGCCATATCCTCTCTCTTCCATGAGAAAGCCGCTTTTGTGTGTAACATGC	1038
QY	1021	TGGCGGTGGGCAAAACATTTGCATATGAAACAAAGCTCATTAGCAATGCTGTGTACATGA	1080
Db	1039	TGGCGGTGGGCAAAACATTTGCATATGAAACAAAGCTCATTAGCAATGCTGTGTACATGA	1098
QY	1081	TCTGTACAGAAGAAATGAAAGCTCAAAAGTCAAAAAATCTCTGAAAAAC-6GAGTTGG	1139
Db	1099	TCTGTACAGAAGAAATGAAAGCTCAAAAGTCAAAAAATCTCTGAAAAACGSGAGCTTTGG	1158
QY	1140	CCTCTCATCTCAGTGGATATATATCCCTTCCCAAAAGAAACAGGCAAGGCTTATCTTTGT	1199
Db	1159	CCTCTCATCTCAGTGGATATATAT-CCCTCCCAAAAGAAACAGGCGCAAGGCTTATCTTTGT	1217
QY	1200	GTCAAAACGGGATTCACCCCACTGTGTGAAAGACAAAGATGCTCTCGACAGTGTGAGTAC	1259
Db	1218	GTCAAAACGGGATTCACCCCACTGTGTGAAAGACAAAGATGCTCTCGACAGTGTGAGTAC	1277
QY	1260	TTACCCCTTGCTAA 1273	
Db	1278	TTACCCCTTGCTAA 1291	
RESULT 3			
AAT14038			
AAT14038 standard: cDNA; 1399 BP			

XX	AC	AAT14038;
XX	DT	07-JUL-1996 (first entry)
XX	DE	Transcription factor-IIIA gene.
XX	KW	Human: transcription factor-IIIA; hTFIIIA; DNA binding protein; ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE; primer: PCR; polymerase chain reaction; foetal brain; anchor primer; diagnostic; probe; transcription control; anticlimour; cancer; therapy; ss.
XX	OS	Homo sapiens.
XX	FH	Key
XX	FT	CDS
XX	FT	Location/Qualifiers 20..1291
XX	FT	/tag= a
XX	FT	/product= Transcription factor-IIIA
XX	FT	/note= "Claim 2"
XX	FT	misc-feature
XX	FT	317..1096
XX	FT	/tag= b
XX	FT	/product= Zinc finger domains
XX	FT	1363..1368
XX	FT	/tag= c
XX	PN	EP704526-A1.
XX	PD	03-APR-1996.
XX	PP	05-SEP-1995; 95EP-0113908.
XX	PR	05-SEP-1994; 94JP-0211022.
XX	PA	(SAKA) OTSUKA PHARM CO LTD.
XX	PI	Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S; WP1: 1996-173033/18. P-PSDB: AAR91305.
XX	DR	
XX	PT	Human Transcription Factor III A gene - useful in regulation of transcription and for diagnosis and treatment of e.g. cancer related diseases
XX	PS	Example 1: Page 11-13; 17pp; English.
XX	CC	The sequence encodes human transcription factor-IIIA (hTFIIIA), a DNA binding protein with 9 zinc finger domains, which is necessary for the initiation of 5S RNA gene transcription, binding to an internal control region of the 5S gene. The coding region (claimed) is given in AAT14037. A fragment lacking a 5'-portion of the gene has been isolated from a human foetal brain cDNA library (ORF7-1), and the 5'-portion of the gene has been isolated by 5'-rapid amplification of cDNA ends using primers H11-R (AAT14039), H11-E (AAT14040), H11-H (AAT14041) and AP-2 (AAT14042), complementary to CC anchor primer AAT14043. Reverse transcription using H-11R is followed by anchor primer ligation and PCR using AP-2 and H11-E, to give a full-length cDNA, ORK7. The gene and its encoded protein may be used in diagnostics, identification or therapy of hereditary diseases such as cancer, or other diseases resulting from abnormal transcriptional control, and to analyse the mechanisms involved in their activity.
XX	SO	Sequence 1399 BP; 405 A; 349 C; 354 G; 291 T; 0 other:
QY	Query Match	96.7%; Score 1231.6; DB 17; Length 1399;
DB	Best Local Similarity	99.5%; Pred. No. 2.5e-311;
	Matches 1267; Conservative	0; Mismatches 4; Indels 3; Gaps 3;
1	ATGCCAGCAGCGGGCGGCGGCCTGTGCTGTCGATACCAGCGCGCTCCGGAAGT	60
20	ATGCCAGCAGCGGGCGGCGGCCTGTGCTGTCGATACCAGCGCGCTCCGGAAGT	79

OY	61	GTGGCGGCGTGGCGCGAAAGTTTATGACGAGGAGCCGTTGGCGCGGCGCGCTTATCCGGC	120
Db	80	GTGGCGGCGTGGCGCGAAAGTTTATGACGAGGAGCCGTTGGCGCGGCGCGCTTATCCGGC	138
OY	121	ACGTGTCTCGGCACGTGGCGAGCGCGCTGTGGCCCTTGGGGCGCGCGGCGCTTGA	180
Db	139	ACGTGTCTCGGCACGTGGCGAGCGCGCTGTGGCCCTTGGGGCGCGCGGCGCTTGA	198
OY	181	TTCGGCGCGCGGTGGTGGCCGAGTCCGGTGTGTCCTTGAACATTCGCGGCGGGTTCATTGC	240
Db	199	TTCGGCGCGCGGTGGTGGCCGAGTCCGGTGTGTCCTTGAACATTCGCGGCGGGTTCATTGC	258
OY	241	AGCGGGGAGAGCTCACACTCCGACCCCGCGCGCGCGCGCTTCCACGAGATTCATCTG	300
Db	259	AGCGGGGAGAGCTCACACTCCGACCCCGCGCGCGCGCGCTTCCACGAGATTCATCTG	318
OY	301	CTCTCTCCCTGACTGTCAGCGCGCAATTACAGCAAAAGCTTGAAGCTTGAACGCGCACTGTG	360
Db	319	CTCTCTCCCTGACTGTCAGCGCGCAATTACAGCAAAAGCTTGAAGCTTGAACGCGCACTGTG	378
OY	361	CAACACACGCGGGGAGAGACATTTCTTTGTATGATGAAGGTGTGGCAAGGCTTCAT	420
Db	379	CAACACACGCGGGGAGAGACATTTCTTTGTATGATGAAGGTGTGGCAAGGCTTCAT	438
OY	421	CAGGAGCTACCATCTGTGAGCGCGCACATTTGTGACTCACAGAGAGAAAAGCGTTTGTGG	480
Db	439	CAGGAGCTACCATCTGTGAGCGCGCACATTTGTGACTCACAGAGAGAAAAGCGTTTGTGG	498
OY	481	TGCGAGCCACTGGCTGTGATCAAAAATTCAACACAAAAATCAAACTTGAAGAAACATTTGA	540
Db	499	TGCGAGCCAAATGGCTGTGATCAAAAATTCAACACAAAAATCAAACTTGAAGAAACATTTGA	558
OY	541	ACGCAAAACATGAAATATCAAAAAACAAATATATATGACTTTTGAAGACTTGAAGAGAC	600
Db	559	ACGCAAAACATGAAATATCAAAAAACAAATATATATGACTTTTGAAGACTTGAAGAGAC	618
OY	601	CTTTAAGAAACATCAGCAGCTGAATAATCCATCAGTGCACGACTACCATGAACCTTATT	660
Db	619	CTTTAAGAAACATCAGCAGCTGAATAATCCATCAGTGCACGACTACCATGAACCTTATT	678
OY	661	CAAGGTATACCAGGAAGAGTGTGGAAACACTTGGTATGATACCACAGCTGAAGACGA	720
Db	679	CAAGGTATACCAGGAAGAGTGTGGAAACACTTGGTATGATACCACAGCTGAAGACGA	738
OY	721	TGCGAAGGCCACAGGGGCTATGTATGTCAAAAAGGATGTTCTTTGTGGCAAAAACATG	780
Db	739	TGCGAAGGCCACAGGGGCTATGTATGTCAAAAAGGATGTTCTTTGTGGCAAAAACATG	798
OY	781	GACGGAACCTTCTGAAACATGTGAGACAAACCCATAAGGAGAAATACATGTGAAAGTATG	840
Db	799	GACGGAACCTTCTGAAACATGTGAGACAAACCCATAAGGAGAAATACATGTGAAAGTATG	858
OY	841	CCGGAAGAAACATTTAAAGCAAGATATTCCTTAAAGCAACATGAAAACTCATGCCCCGA	900
Db	859	CCGGAAGAAACATTTAAAGCAAGATATTCCTTAAAGCAACATGAAAACTCATGCCCCGA	918
OY	901	AAGGATGTATGTGCTGCTGCCAAGAGAGGCTGTGGAGAACCTTACTACTGTGTTTAA	960
Db	919	AAGGATGTATGTGCTGCTGCCAAGAGAGGCTGTGGAGAACCTTACTACTGTGTTTAA	978
OY	961	TCTCCAAAGGCATTTCTCTCTCTTCCATGAGGAAGCGCCCTTTTGTGTGTGAACATGC	1020
Db	979	TCTCCAAAGGCATTTCTCTCTCTTCCATGAGGAAGCGCCCTTTTGTGTGTGAACATGC	1038
OY	1021	TGGCTGTGGCAAAACATTTTGCATGGAAGCAAAAGCTCAGTAGGATGCTGTTGTATATGA	1080
Db	1039	TGGCTGTGGCAAAACATTTTGCATGGAAGCAAAAGCTCAGTAGGATGCTGTTGTATATGA	1098
OY	1081	TCTCTGACAAAGAAATGAAAGCTCAAAACTCAAAAAATCTGCTGA AAAAC - GSGATTTGG	1139
Db	1099	TCTCTGACAAAGAAATGAAAGCTCAAAACTCAAAAAATCTGCTGA AAAACCGGAGTTTGG	1158

QY 1140 CCTCTCATCTCATGTGATATATATCCCTCCCAAGAGAAACAAGGCGCAAGCTTATCTTTGT 1199
 |||||||
 Db 1159 CCTCTCATCTCATGTGATATAT - CCTCCCAAGAGAAACAAGGCGCAAGCTTATCTTTGT 1217
 QY 1200 GTCAAAAGCGAGAGTCACCCCACTGTGTGGAAGACAGATGCTCTGCACACTTGCAGTAC 1259
 |||||||
 Db 1218 GTCAAAAGCGAGAGTCACCCCACTGTGTGGAAGACAGATGCTCTGCACACTTGCAGTAC 1277
 QY 1260 TTACCTTGCTGCTAA 1273
 |||||||
 Db 1278 TTACCTTGCTGCTAA 1291
 RESULT 4
 AAT14037
 ID AAT14037 standard; cDNA; 1269 BP.
 XX
 AC AAT14037:
 XX
 DT 07-JUL-1996 (first entry)
 XX
 DE Transcription factor-IIIA gene.
 XX
 KW Human: transcription factor-IIIA; hTFIIIA; DNA binding protein;
 KW ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;
 KW primer; PCR; polymerase chain reaction; foetal brain;
 KW primer; diagnostic; probe; transcription control;
 KW antitumour; cancer; therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT misc_feature
 FT 336..115
 FT location/qualifiers
 FT /*tag= a
 FT /product= zinc finger domains
 FT 1382..1387
 FT polyA_signal
 FT /*tag= b
 FT
 FT
 PN EP704526-A1.
 XX
 PD 03-APR-1996.
 XX
 PE 05-SEP-1995: 95EP-0113908.
 XX
 PR 05-SEP-1994: 94JP-0211022.
 XX
 PA (SAKA) OTSUKA PHARM CO LTD.
 XX
 PI Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;
 XX
 DR WPI: 1996-173033/18.
 DR P-PSDB: AAR91305.
 XX
 PT Human Transcription Factor III A gene - useful in regulation of
 PT transcription and for diagnosis and treatment of e.g. cancer related
 PT diseases
 XX
 PS
 XX
 CC Claim 2; Page 10-11; 17pp; English.
 CC
 CC The sequence encodes human transcription factor-IIIA (hTFIIIA),
 CC a DNA binding protein with 9 zinc finger domains, which is necessary
 CC for the initiation of 5S RNA gene transcription, binding to an
 CC internal control region of the 5S gene. A fuller cDNA sequence
 CC with flanking regions is given in AAT14038. A fragment lacking a
 CC 5' portion of the gene has been isolated from a human foetal brain
 CC cDNA library (OTK7-1), and the 5'-portion of the gene has been
 CC isolated by 5'-rapid amplification of cDNA ends using primers H11-R
 CC (AAT14039), H11-E (AAT14040), H11-H (AAT14041) and AP-2 (AAT14042),
 CC complementary to anchor primer AAT14043. Reverse transcription using
 CC H-11R is followed by anchor primer ligation and PCR using AP-2 and
 CC H11-E, to give a full-length cDNA, OTK7. The gene and its encoded
 CC protein may be used in diagnosis, identification or therapy of
 CC hereditary diseases such as cancer, or other diseases resulting from

CC abnormal transcriptional control, and to analyse the mechanisms
 CC involved in their activity.
 XX
 SO Sequence 1269 BP; 358 A; 324 C; 329 G; 258 T; 0 other;
 Query Match 96.5%; Score 1228.6; DB 17; Length 1269;
 Best Local Similarity 99.4%; Pred. No. 1.5e-310;
 Matches 1264; Conservative 0; Mismatches 4; Indels 3; Gaps 3;
 QY 1 ATGGCGACGACGGGCGGCGGAGCGGGGCGCTGGTGACCGGGCGGCTCCCGGAGT 60
 |||||||
 Db 1 ATGGCGACGACGGGCGGCGGAGCGGGGCGCTGGTGACCGGGCGGCTCCCGGAGT 60
 QY 61 GTGCCGGCTGCGCGCGAAGTTACAGAGAGAGCCGTGGCGGGCGGG -CGTTCGCCGC 119
 |||||||
 Db 61 GTGCCGGCTGCGCGCGAAGTTACAGAGAGAGCCGTGGCGGGCGGG -CGTTCGCCGC 119
 QY 121 ACGTGTCTGGCAGCTGGGACGCGCGCTGGGCTTGAGAGGCGCGGCTTGA 180
 |||||||
 Db 120 ACGTGTCTGGCAGCTGGGACGCGCGCTGGGCTTGAGAGGCGCGGCTTGA 179
 QY 181 TCCGCGCGCGCTGGTGGCGGAGTGGTGTCTCTTACCATGCCGACGCTTATTC 240
 |||||||
 Db 180 TCCGCGCGCGCTGGTGGCGGAGTGGTGTCTCTTACCATGCCGACGCTTATTC 239
 QY 241 AGCGGCGAGAGCTCAGCTCCGACCGCGCGCGCGCTTCCAGAGTTTCATCTG 300
 |||||||
 Db 240 AGCGGCGAGAGCTCAGCTCCGACCGCGCGCGCGCTTCCAGAGTTTCATCTG 299
 QY 301 CTCTCTCTCAGTGTGACGCGGCAATTACAGCAAAAGCTGGAAGCTTACGCGACCTGTG 360
 |||||||
 Db 300 CTCTCTCTCAGTGTGACGCGGCAATTACAGCAAAAGCTTACGCGACCTGTG 359
 QY 361 CAACGACACGCGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGTGGCAAGGCTTCAT 420
 |||||||
 Db 360 CAACGACACGCGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGTGGCAAGGCTTCAT 419
 QY 421 CAGGAGTACCATCTGAGCCGCGACATTTGCTCTACACAGAGAGAAAGCGTTGTTTG 480
 |||||||
 Db 420 CAGGAGTACCATCTGAGCCGCGCACATTTGCTCTACACAGAGAGAAAGCGTTGTTTG 479
 QY 481 TGCAGCCACTGGCTGTGATCAAAATTCACACAAATCAACTTGAAGAAATTTTGA 540
 |||||||
 Db 480 TGCAGCCAAATGGCTGTGATCAAAATTCACACAAATCAACTTGAAGAAATTTTGA 539
 QY 541 ACGCAACATGAANTCAACAAAGCAATATATATGCAAGTTTGAAGACTTAAAGAGAC 600
 |||||||
 Db 540 ACGCAACATGAANTCAACAAAGCAATATATATGCAAGTTTGAAGACTTAAAGAGAC 599
 QY 601 CTTTAAAGAACATCAGCAGCTGAATAATCCATGCGAGCATACCAATGAACCTTATT 660
 |||||||
 Db 600 CTTTAAAGAACATCAGCAGATGAATAATCCATGCGAGCAATGAACCTTATT 659
 QY 661 CAAGTGTACCCAGAGAGAGATGTGGGAAACACTTGTGCATCACCAGCAAGCTGAAGACGA 720
 |||||||
 Db 660 CAAGTGTACCCAGAGAGATGTGGGAAACACTTGTGCATCACCAGCAAGCTGAAGACGA 719
 QY 721 TGCCAAAGGCCACAGAGGCTGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 780
 |||||||
 Db 720 TGCCAAAGGCCACAGAGGCTGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 779
 QY 781 GACGGAATTTGTGAACATGTGAGAGAAACCATTAAAGAGAAATATATGTGAAGTATG 840
 |||||||
 Db 780 GACGGAATTTGTGAACATGTGAGAGAAACCATTAAAGAGAAATATATGTGAAGTATG 839
 QY 841 CCGGAAAAATTTAAACGCAAAAGATTACCTTAAGCAACATGAAACTCATGCCCCAGA 900
 |||||||
 Db 840 CCGGAAAAATTTAAACGCAAAAGATTACCTTAAGCAACATGAAACTCATGCCCCAGA 899
 QY 901 AAGGATGTATGTGCTGTCTCCAAAGAAAGCTGTGGAAGAACTATATCTGTGTTTAA 960
 |||||||
 Db 900 AAGGATGTATGTGCTGTCTCCAAAGAAAGCTGTGGAAGAACTATATCTGTGTTTAA 959

QY 961 TCCTCAAGCCATATCCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGAACATGC 1020
|||||
Db 960 TCTCCAAAGCATATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGAACATGC 1019
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGCTCTCAGTACGCTGTTGTACATGA 1080
|||||
Db 1020 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGCTCTCAGTACGCTGTTGTACATGA 1079
QY 1081 TCCGACAGCAAGAAATGAAGCTCAAGTCAAAATCTCGTGAAGAAC-GGAGTTTGG 1139
|||||
Db 1080 TCCGACAGCAAGAAATGAAGCTCAAGTCAAAATCTCGTGAAGAACGGAGTTTGG 1139
QY 1140 CCTCTCATCTCAGTGAATATATCCCTCCCAAGAAAGCAAGGCAAGCTTATCTTTGT 1199
|||||
Db 1140 CCTCTCATCTCAGTGAATATAT-CCCTCCAAAGAAAGCAAGGCAAGCTTATCTTTGT 1198
QY 1200 GTCAAAAGGAGAGTCAACCACTGTGTGAGACAGCAAGATGCTCTGACAGTTGCACTAC 1259
|||||
Db 1199 GTCAAAAGGAGAGTCAACCACTGTGTGAGACAGCAAGATGCTCTGACAGTTGCACTAC 1258
QY 1260 TTACCTTTGGC 1270
|||||
Db 1259 TTACCTTTGGC 1269

RESULT 5
AAAI5407
ID AAI5407 standard; DNA: 1213 BP.

AC AAI5407;
DT 04-SEP-2000 (first entry)
XX

DE Fragment of DNA encoding a transcription factor designated htfIIIA.

KM Human: transcription factor: htfIIIA; DNA-binding protein;
KW transcription; ribosomal RNA 5S gene; transcriptional control;
XX cancer; ss.
XX
OS Homo sapiens.

PN MO200028024-A1.

PD 18-MAY-2000.

PF 09-NOV-1999; 99MO-FR02738.

PR 10-NOV-1998; 98FR-0014146.

PA (HMRI) HOECHST MARION ROUSSEL.

PI Bordon-Pallier F, Rocher C;

DR WPI; 2000-387419/33.

PT New nucleic acid encoding human transcription factor IIA, useful for
treatment and diagnosis of cancer and inherited disease --

PS Claim 4; Page 42; 49pp; French.

XX The present sequence represents a fragment of a human transcription
CC factor (designated htfIIIA) gene. The polypeptide is probably a
CC DNA-binding protein involved in initiating transcription of
CC the gene for ribosomal RNA 5S and maintaining the stability of
CC transcription of other control genes. The htfIIIA polynucleotides and
CC polypeptides are used to make therapeutic or diagnostic compositions
CC for diseases associated with disorders of transcriptional control,
CC particularly cancer or other inherited diseases. The htfIIIA
CC polynucleotide can also be used to detect anomalies in gene
CC transcription, particularly for diagnosis of inherited disease, also
CC for studying diseases involving htfIIIA.

XX Sequence 1213 BP; 349 A; 309 C; 302 G; 253 T; 0 other;

Query Match 95.2%; Score 1211.4; DB 21; Length 1213;
Best Local Similarity 99.9%; Pred. No. 4.5e-306;
Matches 1212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 GTGCGGCGCTGCGCGCGAAGGTTGAGCAGGAGCGCGCGCGCGCGCGGTTCCCGGC 120
|||||
Db 1 GTGCGGCGCGCGCGCGAAGGTTGAGCAGGAGCGCGCGCGCGCGCGGTTCCCGGC 60
QY 121 ACGTGTCTGGCAGCTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
|||||
Db 61 ACGTGTCTGGCAGCTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 181 TCCGCGCGCGCGTGGTGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240
|||||
Db 121 TCCGCGCGCGCGTGGTGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 241 AGCGGCGGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
|||||
Db 181 AGCGGCGGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 301 CTGCTTCCCTGACTGCGAGCGCGCAATTACAGCAAAAGCTTGAAGCTTACCGCCTGTG 360
|||||
Db 241 CTGCTTCCCTGACTGCGAGCGCGCAATTACAGCAAAAGCTTGAAGCTTACCGCCTGTG 300
QY 361 CAAGCACACGGGGAGAGACCATTTGTTGACTATGAAGGGGTGGCAAGGCGCTTCAT 420
|||||
Db 301 CAAGCACACGGGGAGAGACCATTTGTTGACTATGAAGGGGTGGCAAGGCGCTTCAT 360
QY 421 CAGGAGCTACCATCTGAGCGCGCCACATTCGACTCACAGAGAGAAAGCCGTTTGTG 480
|||||
Db 361 CAGGAGCTACCATCTGAGCGCGCCACATTCGACTCACAGAGAGAAAGCCGTTTGTG 420
QY 481 TGCAGCCACTGGCTGTATATATAAATTCACACAAATTCAACTTGAAGAACTTTTGA 540
|||||
Db 421 TGCAGCCACTGGCTGTATATATAAATTCACACAAATTCAACTTGAAGAACTTTTGA 480
QY 541 ACGCAACATGAAATATCAACAAACAAATATATATGCAATTTGAAAGCTTGAAGAGC 600
|||||
Db 481 ACGCAACATGAAATATCAACAAACAAATATATATGCAATTTGAAAGCTTGAAGAGC 540
QY 601 CTTTAAAGAAACATCAGAGCTGAAATATCATGCTGCACGATACCAATGAACTCTATT 660
|||||
Db 541 CTTTAAAGAAACATCAGAGCTGAAATATCATGCTGCACGATACCAATGAACTCTATT 600
QY 661 CAACTGTACCCAGGAAAGATGTGGAAACACTTTCATCAACGCAAGCTGAAAGCACA 660
|||||
Db 601 CAACTGTACCCAGGAAAGATGTGGAAACACTTTCATCAACGCAAGCTGAAAGCACA 660
QY 721 TGCAGAGCCAGAGGAGCTATGATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 780
|||||
Db 661 TGCAGAGCCAGAGGAGCTATGATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 720
QY 781 GACGGAACCTTCTGAACATGTGAGAGAAACCATTAAGAGAAATATATGTGAAGTATG 840
|||||
Db 721 GACGGAACCTTCTGAACATGTGAGAGAAACCATTAAGAGAAATATATGTGAAGTATG 840
QY 841 CCGGAAACATTTAAACGCAAGATTACCTTAAGCAACATGAAACTATGCCCCAGA 900
|||||
Db 781 CCGGAAACATTTAAACGCAAGATTACCTTAAGCAACATGAAACTATGCCCCAGA 840
QY 901 AAGGATGTATGTGCTGCTCAAGAGAGCTGTGGAAGAACTATCTACTGTGTTAA 960
|||||
Db 841 AAGGATGTATGTGCTGCTCAAGAGAGCTGTGGAAGAACTATCTACTGTGTTAA 900
QY 961 TCTCCAAAGCATATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGAACATGC 1020
|||||
Db 901 TCTCCAAAGCATATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGAACATGC 960
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGCTCTCAGTACGCTGTTGTACATGA 1080
|||||
Db 961 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGCTCTCAGTACGCTGTTGTACATGA 1020

QY 1081 TCCTGACAAAGAAAGAAATGAAGCTCAAGTCAAAAAATCTCTGTAAGAAACGAGATTGCG 1140
|||||
Db 1021 TCCTGACAAAGAAAGAAATGAAGCTCAAGTCAAAAAATCTCTGTAAGAAACGAGATTGCG 1080
QY 1141 CTCATCTCATGATGATATATCTCCCAAAAGCAAGGCAAGCTTATCTTTGTG 1200
|||||
Db 1081 CTCATCTCATGATGATATATCTCCCAAAAGCAAGGCAAGCTTATCTTTGTG 1140
QY 1201 TCAAAAGCAAGTCAAGCAAGTCTGTGGAAGACAGATGCTCTGACAGTTGCACTACT 1260
|||||
Db 1141 TCAAAAGCAAGTCAAGCAAGTCTGTGGAAGACAGATGCTCTGACAGTTGCACTACT 1200
QY 1261 TACCTTGGCTAA 1273
|||||
Db 1201 TACCTTGGCTAA 1213

RESULT 6
AAH34586
ID AAH34586 standard; cDNA: 1349 BP.
XX
AC AAH34586;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1668.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 13; ss.
XX
OS Homo sapiens.
XX
PN MO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Blirise CE, Rosen CA;
DR WPI: 2001-235357/24.
DR P-PSDB: AAG75181.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3273-3274; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX

50 Sequence 1349 BP; 400 A; 337 C; 325 G; 285 T; 2 other;
Query Match 94.7%; Score 1206; DB 22; Length 1349;
Best Local Similarity 99.28; Pred. No. 1.2e-304;
Matches 1212; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 52 CCCGAAAGTGTGGCGGGCTCGCGGAAAGTTTCACACAGGAGCCGTGGGCGGGCGCG 111
||| |
Db 11 CCCAGCGCTCGGGGGCGCGCGCGGAAAGTTTCACACAGGAGCCGTGGGCGGGCGCG 70
QY 112 GTTCCCGGACGTGTCTTCGACGTGGCAGCGCGCTGGCCCTGGGCTTGGAGGCGCGG 171
|||
Db 71 GTTCCCGGACGTGTCTTCGACGTGGCAGCGCGCTGGCCCTGGGCTTGGAGGCGCGG 130
QY 172 CGCCCTGGATCGCGCGCGCGCGGTCCCGAGTCCGTGCTGCTTGAACATCGCGGACGC 231
|||||
Db 131 CGCCCTGGATCGCGCGCGCGCGGTCCCGAGTCCGTGCTGCTTGAACATCGCGGACGC 190
QY 232 GTTCATTGACAGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCTTCCAGAGAG 291
|||||
Db 191 GTTCATTGACAGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCTTCCAGAGAG 250
QY 292 GTTCATCTGCTCTTCCCTGACGTGCAGCGCCCAATTACAGCAAGCCCTGGAAGCTTGAAGC 351
|||||
Db 251 GTTCATCTGCTCTTCCCTGACGTGCAGCGCCCAATTACAGCAAGCCCTGGAAGCTTGAAGC 310
QY 352 GCACCTGTGCAGACAGACAGAGGGGAGAGACATTTTGTGTACTATGAAAGGTGTGGCAA 411
|||||
Db 311 GCACCTGTGCAGACAGACAGAGGGGAGAGACATTTTGTGTACTATGAAAGGTGTGGCAA 370
QY 412 GGCCCTTCATCAGGGGACTACATCTGAGCGCGCACATTTCTGACTACACAGAGAGAAAGCC 471
|||||
Db 371 GGCCCTTCATCAGGGGACTACATCTGAGCGCGCACATTTCTGACTACACAGAGAGAAAGCC 430
QY 472 GTTGTGTTGTGACGACCTGCGCTGTGATCAAAATTCACACAAATTCAACTTGAAGAA 531
|||||
Db 431 GTTGTGTTGTGACGACCTGCGCTGTGATCAAAATTCACACAAATTCAACTTGAAGAA 490
QY 532 ACATTTGGAAGCAACATCTGAAATCAACCAAAACAAATATATATGCAATTTTGAAGACTG 591
|||||
Db 491 ACATTTGGAAGCAACATCTGAAATCAACCAAAACAAATATATATGCAATTTTGAAGACTG 550
QY 592 TAAGAAAGACTTTTAAAGAAATCATCAGACAGTGAATAATCCATGAGTGCATATACCAATGA 651
|||||
Db 551 TAAGAAAGACTTTTAAAGAAATCATCAGACAGTGAATAATCCATGAGTGCATATACCAATGA 610
QY 652 ACCTCTATTCAAGTGTACCCAGGAAGGATGTGGAAACACTTTTGCATCCACCAAGCT 711
|||||
Db 611 ACCTCTATTCAAGTGTACCCAGGAAGGATGTGGAAACACTTTTGCATCCACCAAGCT 670
QY 712 GAACGACATGCCCAAGGCGCCACAGAGGCTATGTATGTAAAGAGATGTTCTTTGTGGC 771
|||||
Db 671 GAACGACATGCCCAAGGCGCCACAGAGGCTATGTATGTAAAGAGATGTTCTTTGTGGC 730
QY 772 AAAAAACATGACGAGCACTTCTGAAAAATGTAGAGAAACCATTAAGAGAAATACATCT 831
|||||
Db 731 AAAAAACATGACGAGCACTTCTGAAAAATGTAGAGAAACCATTAAGAGAAATACATCT 790
QY 832 TGAAGTATGCCGGAAGCAATTTAAACGCAAGATTTACCTTAAGCAACATGAAGAACTCA 891
|||||
Db 891 TGAAGTATGCCGGAAGCAATTTAAACGCAAGATTTACCTTAAGCAACATGAAGAACTCA 850
QY 892 TGCCCCCAAGAGGATGTATGTGCTGTCCCAAGAGAAAGGCTGTGAAGAACCTTATACAC 951
|||||
Db 851 TGCCCCCAAGAGGATGTATGTGCTGTCCCAAGAGAAAGGCTGTGAAGAACCTTATACAC 910
QY 952 TGTGTTTAATCTCCAAAGCAATATCTCTCTTCATGAGAGAAAGCCGCTTTTGTGTG 1011
|||||
Db 911 TGTGTTTAATCTCCAAAGCAATATCTCTCTCTTCATGAGAGAAAGCCGCTTTTGTGTG 970
QY 1012 TGAACATCTGCTGTGTGGCAAAACATTTGCAATGAAACAAAGTCTCACTAGACATGCTGT 1071
|||||
Db 971 TGAACATCTGCTGTGTGGCAAAACATTTGCAATGAAACAAAGTCTCACTAGACATGCTGT 1030

QY 1072 TGTACATGATCTGCACAGAGAAAGTCAAGTCAAAATCTCGTGAACG 1131
|||||
Db 1031 TGTACATGATCTGCACAGAGAAAGTCAAGTCAAAATCTCGTGAACG 1090
QY 1132 GAGTTGGCCTTCATCTCAGTGTATATCCCTCCCAAGAGGCAAGGCTT 1191
|||||
Db 1091 GAGTTGGCCTTCATCTCAGTGTATATCCCTCCCAAGAGGCAAGGCTT 1150
QY 1192 ATCTTGTGCAAAAGGAGAGTCACTGTGTGAAGACAGATGCTCGACAGT 1251
|||||
Db 1151 ATCTTGTGCAAAAGGAGAGTCACTGTGTGAAGACAGATGCTCGACAGT 1210
QY 1252 TGCAGTACTTACCTTGCTTAA 1273
|||||
Db 1211 TGCAGTACTTACCTTGCTTAA 1232

RESULT 7
AB054923/c
ID AB054923 standard; cDNA; 690 BP.
XX
AC AB054923;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HUSYA18 cDNA, SEQ ID NO:803.
XX

Human: ovarian antigen; ovary; ovarian; breast; cancer; tumour;
ovarian cancer; breast cancer; tumour; reproductive system disorder;
infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
inflammatory condition; immune disorder; blood disorder;
cardiovascular disorder; respiratory disorder; neurological disorder;
gastrointestinal disorder; urinary system disorder; drug screening;
gene therapy; chromosome mapping; forensic analysis;
antibody preparation; cytostatic; immunomodulatory; neuroprotective;
antiinflammatory; gynaecological; reproductive; chromosome 13q12.3-13.1;
gene; ss.
XX
OS Homo sapiens.
XX
PN WC02020677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
DR P-PSDB; ABP41846.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 1; SEQ ID No 803; 2922pp; English.
XX

The invention relates to 2175 novel human ovarian antigens (ABP41054-
ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and

CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 690 BP; 154 A; 143 C; 145 G; 243 T; 5 other:
XX

Query Match 38.9%; Score 495; DB 24; Length 690;
Best Local Similarity 94.8%; Pred. No. 3,9e-119;
Matches 529; Conservative 2; Mismatches 25; Indels 2; Gaps 2;

QY 716 CGACATGCCAAGGCCCAAGGCGTATGATGTCAAAAAGAGTTCCTTTGGCAAA 775
|||
Db 690 CGCAGTCCCAAGGCCCAAGGCGTATGATGTC-AAGAGATGTCCTTTGGTGC-AA 633
QY 776 ACATGAGCGGACCTTCTGAAACATGTGAGAGAAACCATTAAGGAAATACATCTGAA 835
|||||
Db 632 ACATGAGCGGACCTTCTGAAACATGTGAGAGAAACCATTAAGGAAATACATCTGAA 573
QY 836 GTATGCGGGAACATTTAAACGCAAGATTAACCTTAAGCAACATGAATCATATGCC 895
572 GTATGCGGGAACATTTAAACGCAAGATTAACCTTAAGCAACATGAATCATATGCC 513
Db 896 CCAAGAAAGGATGTATGTGCTGTCAGAGAGAGGCTGTGAGAACCTATACATCTG 955
|||||
QY 512 CAGAAAGGATGTATGTGCTGTCAGAGAGAGGCTGTGAGAACCTATACATCTG 453
|||||
QY 956 TTTATCTCCAAAGCATATCTCTCCATAGAGAAAGCCGCTTTGTTGTGAA 1015
|||||
Db 452 TTTATCTCCAAAGCATATCTCTCTCATATAGAGAAAGCCGCTTTGTTGTGAA 393
QY 1016 CATGCTGGCTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGCATGTGTGTA 1075
|||||
Db 392 CATGCTGGCTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGCATGTGTGTA 333
QY 1076 CATGATCTCTGACAGAGAAATTAAGCTCAAAAGTCTCTGTAAGAAAGGAGT 1135
|||||
Db 332 CATGATCTCTGACAGAGAAATTAAGCTCAAAAGTCTCTGTAAGAAAGGAGT 273
QY 1136 TTGGCTCTCATCTCAGTGTATATCCCTCCCAAGAGGAGGCAAGGCTTATCT 1195
|||||
Db 272 TTGGCTCTCATCTCAGTGTATATCCCTCCCAAGAGGAGGCAAGGCTTATCT 213
QY 1196 TTGTGTCAAAAGGAGAGTCAACCACTGTGTGAAGACAGATGCTCTGACAGTTGCA 1255
|||||
Db 212 TTGTGTCAAAAGGAGAGTCAACCACTGTGTGAAGACAGATGCTCTGACAGTTGCA 153
QY 1256 GTACTTACCTTGCTTAA 1273
|||||
Db 152 GTACTTACCTTGCTTAA 135

RESULT 8
ABK44775/c
ID ABK44775 standard; cDNA; 439 BP.
XX

AC ABK4775;
XX
DT 05-JUN-2002 (first entry)
XX
DE cDNA encoding colon tumour protein, SEQ ID No 326.
XX
KM Human; colon tumour; vaccine; colon cancer; immunogenic;
XX immunotherapy; gene; ss.
OS Homo sapiens.
XX
PN WO200212328-A2.
XX
PD 14-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US24218.
XX
PR 03-AUG-2000; 2000US-223283P.
XX 28-MAR-2001; 2001US-279763P.
XX 29-JUN-2001; 2001US-302051P.
XX
XX (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secrist H;
XX WPI; 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
XX for diagnosing, preventing, and treating colon cancer, and as markers
XX for the progression of cancer -
XX
PS Claim 1; SEQ ID No 326; 147pp: English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 439 BP; 106 A; 94 C; 99 G; 140 T; 0 other;
XX
Query Match 32.9%; Score 418.4; DB 24; Length 439;
Best Local Similarity 99.8%; Pred. No. 3.2e-99;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 854 AAACGCAAGATTACTTAAACCAACATGAAACTGATGCCCGAAGAGGATGTATGT 913
DB 439 AAAGCGAAAGATTACTTAAACCAACATGAAACTGATGCCCGAAGAGGATGTATGT 380
XX
QY 914 CGCTGTCCAGAGAGGCTGTGGAAGAACCTATACTGTGTATCTCCAAAGCCAT 973
DB 379 CGCTGTCCAGAGAGGCTGTGGAAGAACCTATACTGTGTATCTCCAAAGCCAT 320
XX
QY 974 ATTCGTCCCTTCATGAGGAAGCGCCCTTTGTGTGTAACATGCGTGGTGGCAAA 1033
DB 319 ATTCGTCCCTTCATGAGGAAGCGCCCTTTGTGTGTAACATGCGTGGTGGCAAA 260
XX
QY 1034 ACATTTGCAATGAACAAGTCTCAGTAGGCATGCTGTGTACATGATCCTGACAAGAG 1093
DB 259 ACATTTGCAATGAACAAGTCTCAGTAGGCATGCTGTGTACATGATCCTGACAAGAG 200
XX
QY 1094 AAAATGAAGCTCAAAAGTCAAAATCTCTGTGTAAGAAACGAGATTTGGCCTCTCATCTCAGT 1153
DB 199 AAAATGAAGCTCAAAAGTCAAAATCTCTGTGTAAGAAACGAGATTTGGCCTCTCATCTCAGT 140

QY 1154 GGATATATCCCTCCCAAAAGGAACAGGCTTATCTTTGTGTCAAAACGAGAG 1213
DB 139 GGATATATCCCTCCCAAAAGGAACAGGCTTATCTTTGTGTCAAAACGAGAG 80
XX
QY 1214 TCACCAACTGTGTGGAAGACAAGATGCTCTGCACAGTTGACATCTTGGCTAA 1273
DB 79 TCACCAACTGTGTGGAAGACAAGATGCTCTGCACAGTTGACATCTTGGCTAA 20
XX
RESULT 9
ID ABK45138 standard; cDNA; 439 BP.
XX
XX ABK45138;
XX
AC ABK45138;
XX
DT 05-JUN-2002 (first entry)
XX
DE cDNA encoding colon tumour protein, SEQ ID No 689.
XX
XX
KM Human; colon tumour; vaccine; colon cancer; immunogenic;
XX immunotherapy; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200212328-A2.
XX
PN 14-FEB-2002.
XX
PD 31-JUL-2001; 2001WO-US24218.
XX
PF 03-AUG-2000; 2000US-223283P.
XX 28-MAR-2001; 2001US-279763P.
XX 29-JUN-2001; 2001US-302051P.
XX
XX (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secrist H;
XX WPI; 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
XX for diagnosing, preventing, and treating colon cancer, and as markers
XX for the progression of cancer -
XX
PS Claim 1; SEQ ID No 689; 147pp: English.
XX
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 439 BP; 141 A; 99 C; 93 G; 106 T; 0 other;
XX
Query Match 32.7%; Score 416.8; DB 24; Length 439;
Best Local Similarity 99.5%; Pred. No. 8.3e-99;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 854 AAACGCAAGATTACTTAAACCAACATGAAACTGATGCCCGAAGAGGATGTATGT 913
DB 1 AAACGCAAGATTACTTAAACCAACATGAAACTGATGCCCGAAGAGGATGTATGT 60
XX
QY 914 CGCTGTCCAGAGAGGCTGTGGAAGAACCTATACTGTGTATCTCCAAAGCCAT 973

```

|||||
Db 61 CGCTGTCCAGACAGAGCTGTGGAAAGCACTATACAACTGTGTTATATCTCCAAAGCCAT 120
Oy 974 ATCCCTCTCCATGAGGAAAGCGCCCTTTGTGTGAACATGCGTGGCGCAA 1033
Db 121 ATCTCTCTCTCCATGAGGAAAGCGCCCTTTGTGTGAACATGCGTGGCGCAA 180
Oy 1034 ACATTTGCAATGAAACAAAGCTCTACTAGGCATGCTTGTACATGATGCTGACAAAG 1093
Db 181 ACATTTGCAATGAAACAAAGCTCTACTAGGCATGCTTGTACATGATGCTGACAAAG 240
Oy 1094 AAAATGAAGCTCAAAAGTCAAAAATCTCTGTGAAAAACGAGTTGGCCTCTCATCTAGT 1153
Db 241 AAAATGAAGCTCAAAAGTCAAAAATCTCTGTGAAAAACGAGTTGGCCTCTCATCTAGT 300
Oy 1154 GATATATCTCCCTCCCAAAAGGAAACAGGCAAGGCTTATCTTGTGTCAAAACGAGAG 1213
Db 301 GATATATCTCCCTCCCAAAAGGAAACAGGCAAGGCTTATCTTGTGTCAAAACGAGAG 360
Oy 1214 TCACCCAACCTGTGTGAAGACAGAGATGCTCTGCACAGTTGCTACTTACCTTTGCGTAA 1273
Db 361 TCACCCAACCTGTGTGAAGACAGAGATGCTCTGCACAGTTGCTACTTACCTTTGCGTAA 420

RESULT 10
ABO59499
ID ABO59499 standard; cDNA: 537 BP.
XX
AC ABO59499:
XX
DT 02-AUG-2002 (first entry)
DE Human colon cancer related nucleotide sequence SEQ ID NO:3194.
XX
KM Human: colon cancer; cancer: tissue profiling; forensic; mapping;
KM genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN NC0200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCY-2001; 2001WO-US30732.
XX
PR 02-OCY-2000; 2000US-237271P.
XX
PA (FARB ) BAYER CORP.
XX
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thleglingam A, Lewis ME;
XX
DR WPI: 2002-426115/45.
XX
PT New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell
PT or tissue type, and in antisense therapy -
XX
PS Claim 1; Fig 1; 796pp: English.
XX
CC ABO56306 to ABO60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC encoded by the ABO60776 to ABO60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridizes to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence of
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the

```

```

CC corresponding gene residues, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists.
XX
SO Sequence 537 BP; 196 A; 116 C; 106 G; 117 T; 2 other:
Query Match 29.4%; Score 374; DB 24; Length 537;
Best Local Similarity 96.6%; Pred. No. 1,4e-87;
Matches 402; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
Oy 662 AAGGTATCCCAAGAGATGTGGGAAACACTTTGTCATCACCAGCAAGCTGAAACGACAT 721
Db 123 AAAAGGTCCCAAGAGATGTGGGAAACACTTTGTCATCACCAGCAAGCTGAAACGACAT 182
Oy 722 GCCAAGGCCCAAGAGGCTATGTATGTCAAAAGAGATGTCCTTTGTGGCAAAAACATGG 781
Db 183 GCCAAGGCCCAAGAGGCTATGTATGTCAAAAGAGATGTCCTTTGTGGCAAAAACATGG 242
Oy 782 ACCGAACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGC 841
Db 243 ACGGAATCTTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATGC 302
Oy 842 CGGAAAAATTTTAAACCAAGATTACCTTACGACACATGAAAACTGATGCCCAAGAA 901
Db 303 CGGAAAAATTTTAAACCAAGATTACCTTACGACACATGAAAACTGATGCCCAAGAA 362
Oy 902 AGGATGTATGTGCGCTGTCGAAGAGAGCGTGTGGAAGACCTATACTAGTGTGTTAAT 961
Db 363 AGGATGTATGTGCGCTGTCGAAGAGAGCGTGTGGAAGACCTATACTAGTGTGTTAAT 421
Oy 962 CTCCAAGCCATATCTCTCTCCATGAGGAAAGCCGCCCTTTGTGTGTGAACATGCT 1021
Db 422 CTCCAAGCCATATCTCTCTCCATGAGGAAAGCCGCCCTTTGTGTGTGAACATGCT 481
Oy 1022 GCGTGTGGCAAAACA-TTTCGATGAAACAAAGTCTACTAGGATCTGTTGATC 1076
Db 482 GCGTGTGGCAAAACA-TTTCGATGAAACAAAGTCTACTAGGATCTGTTGATC 537

RESULT 11
ABV49231
ID ABV49231 standard; cDNA: 466 BP.
XX
AC ABV49231:
XX
DT 17-SEP-2002 (first entry)
DE Human prostate expression marker cDNA 49222.
XX
KM Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN NC0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX

```

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 9624; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 466 BP; 153 A; 92 C; 102 G; 119 T; 0 other:

Query Match 23.7%; Score 301.4; DB 23; Length 466;

Best Local Similarity 99.7%; Pred. No. 1.2e-68;

Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 737 GGCATGATATCAAAAAGCATGTTCTTGTGGCAAAAACATGAGCGAATCTTGAAA 796

DB 164 GGCATGATATCAAAAAGCATGTTCTTGTGGCAAAAACATGAGCGAATCTTGAAA 223

QY 797 CATGTGAGAGAAAGCCATTAAGAGAAATATCTATGTGAGTATGCCGAGAAATTTTAA 856

DB 224 CATGTGAGAGAAAGCCATTAAGAGAAATATCTATGTGAGTATGCCGAGAAATTTTAA 283

QY 857 CGCAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGAAAGGATGTATGTCG 916

DB 284 CGCAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGAAAGGATGTATGTCG 343

QY 917 TGTCCAGAGAAAGGCTGTGGAAGCACTATCTACTGTGTTAATCTCCAAAGCCATATC 976

DB 344 TGTCCAGAGAAAGGCTGTGGAAGCACTATCTACTGTGTTAATCTCCAAAGCCATATC 403

QY 977 CTCCTCCATTCATGAGAAAGCGCCCTTTGTGTGTAACATGCTGTGGGCAAAACA 1036

DB 404 CTCCTCCATTCATGAGAAAGCGCCCTTTGTGTGTAACATGCTGTGGGCAAAACA 463

QY 1037 TTT 1039

DB 464 TTT 466

RESULT 12

ABV19459

ID ABV19459 standard; cDNA: 425 BP.

XX AC ABV19459;

XX 13-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 19450.

XX

KW Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

OS

XX WO200160860-A2.

PN

XX 23-AUG-2001.

PD

XX

PF 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Endege WO, Monahan JE;

XX

DR WPI: 2001-662795/76.

XX

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 3184; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 425 BP; 142 A; 79 C; 90 G; 114 T; 0 other:

Query Match 23.6%; Score 300.4; DB 23; Length 425;

Best Local Similarity 99.7%; Pred. No. 2e-68;

Matches 301; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 737 GGCATGATATCAAAAAGCATGTTCTTGTGGCAAAAACATGAGCGAATCTTGAAA 796

DB 124 GGCATGATATCAAAAAGCATGTTCTTGTGGCAAAAACATGAGCGAATCTTGAAA 183

QY 797 CATGTGAGAGAAAGCCATTAAGAGAAATATCTATGTGAGTATGCCGAGAAATTTTAA 856

DB 184 CATGTGAGAGAAAGCCATTAAGAGAAATATCTATGTGAGTATGCCGAGAAATTTTAA 243

QY 857 CGCAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGAAAGGATGTATGTCG 916

DB 244 CGCAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGAAAGGATGTATGTCG 303

QY 917 TGTCCAGAGAAAGGCTGTGGAAGCACTATCTACTGTGTTAATCTCCAAAGCCATATC 976

DB 304 TGTCCAGAGAAAGGCTGTGGAAGCACTATCTACTGTGTTAATCTCCAAAGCCATATC 363

QY 977 CTCCTCCATTCATGAGAAAGCGCCCTTTGTGTGTAACATGCTGTGGGCAAAACA 1036

DB 364 CTCCTCCATTCATGAGAAAGCGCCCTTTGTGTGTAACATGCTGTGGGCAAAACA 423

QY 1037 TT 1038

DB 424 TT 425

RESULT 13

AAH22129/C

ID AAH22129 standard; cDNA: 1401 BP.

XX AC AAH22129;

XX

DT 20-AUG-2001 (first entry)
XX Human eukaryotic Initiation factor 3 (heIF3) encoding cDNA SEQ ID NO:6.
DE
XX
KW Human: eIF3; eukaryotic Initiation factor; translation Initiation factor;
KW heIF3; dendritic cell; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 237..1073
FT /"cag" a
FT /product= "eukaryotic Initiation factor (heIF3)"
PN CNI281044-A.
XX
PD 24-JAN-2001.
XX
PF 27-JUN-2000; 2000CN-0116791.
XX
PR 27-JUN-2000; 2000CN-0116791.
XX
PI (NANF-) NANFANG RES CENT STATE HUMAN GENE GROUP.
XX
PI Yang Y, Xiao H, Kang B;
XX
DR WPI: 2001-282654/30.
DR P-PSDB: AAB98214.
XX
PT Human translation Initiation factor protein and its coding sequence -
PS
PS Claim 1: Page 17 (disclosure): 20pp; Chinese.
XX
CC The present invention describes a human eukaryotic Initiation factor
CC (eIF3), designated heIF3, which is expressed in human dendritic cells.
CC heIF3 has translation Initiation factor activity. Also described in the
CC present invention are methods for the preparation and detection of the
CC heIF3 protein and nucleotide sequences. The present sequence encodes
CC heIF3, as given in the present invention.
XX
SQ Sequence 1401 BP; 459 A; 281 C; 306 G; 355 T; 0 other;
Query Match 19.38; Score 246; DB 22; Length 1401;
Best Local Similarity 100.0%; Pred. No. 5.7e-54;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1028 GGCAAAACATTGGCAATGAACAAAGTCTCAGTGGATGCTGTTGACATGATCCTGAC 1087
DB 1401 GGCAAAACATTGGCAATGAACAAAGTCTCAGTGGATGCTGTTGACATGATCCTGAC 1342
QY 1088 AAGAAGAAATGAAGCTCAAAAGTCAAAATCTCGTGAAGAAAGAGAGTTTGGCCTTCAT 1147
DB 1341 AAGAAGAAATGAAGCTCAAAAGTCAAAATCTCGTGAAGAAAGAGAGTTTGGCCTTCAT 1282
QY 1148 CTCAGTGTATATTCCTCCCAAAAGCAAGGAGCGACTTATCTTTGTGCAAAAC 1207
DB 1281 CTCAGTGTATATTCCTCCCAAAAGCAAGGAGCGACTTATCTTTGTGCAAAAC 1222
QY 1208 GCAAGTACACCACTGTGTGGAAGACAAGATGCTCGACAGTTGACATGACTTACCCCT 1267
DB 1221 GCAAGTACACCACTGTGTGGAAGACAAGATGCTCGACAGTTGACATGACTTACCCCT 1162
QY 1268 GGCTTAA 1273
DB 1161 GGCTTAA 1156

RESULT 14
AAS73534
ID AAS73534 standard; cDNA; 1889 BP.
XX
AC AAS73534;
XX

DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #9338.
DE
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI DermaNac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG09347.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1: SEQ ID NO 9338; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at http://wipo.int/pub/published_pat_sequences.
SQ Sequence 1889 BP; 488 A; 453 C; 417 G; 531 T; 0 other;
Query Match 18.58; Score 236; DB 23; Length 1889;
Best Local Similarity 95.58%; Pred. No. 2.7e-51;
Matches 275; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
QY 986 CATAGAGAAAGCCGCCCTTTGTGTGTGACACATGCTGTGGCAAAACATTGGCAATG 1045
DB 2 CATAGAGAAAGCCGCCCTTTGTGTGTGACACATG-TGGCTGTGGCAAAACA-TTGGCAAT 59
QY 1046 AAACAAAGTCTACATAGGCAATGCTGTGTATCATATCTGACAGAAAGAAATGACCTC 1105
DB 60 AAACCAAGTCTACATAGGCAATGCTGTGTATCATG-TCTGTGCAAGAAAGAAATGACCTC 118
QY 1106 AAAGTCAAAAATCTCGTGAAGAAAGAGATTGGCCTTCATCTCAGTGAATATATCCCT 1165
DB 119 AAAGTCAAAAATCTCGTGAAGAAAGAGATTGGCCTTCATCTCAGTGAATATATCCCT 178
QY 1166 CCCAAAAGAAACAAAGGCAAGGCTTATCTTTGTGTCAAAAAGAGAGACTCACCAACTGT 1225

THIS PAGE BLANK (USPTO)

QY	361	CAACACACGGGGAGAGACATTTTGTGTCACATGAAGGTTGGCAAGCCCTTAT	420
Db	380	CAGACACAGGGGGAGAGACATTTGTGTGACATGTGAAGGTTGGCAAGCCCTTAT	439
QY	421	CAGGACATACATCTGTGAGCCGCCACATTTGCATCACAGGAGAAAAGCCGTTGTTG	480
Db	440	CAGGACATACATCTGTGAGCCGCCACATTTGCATCACAGGAGAAAAGCCGTTGTTG	499
QY	481	TGCGAGCCACTGGCTGTGATATAAAAATTCAACACAAATTAACCTTGAAGAACATTTTGA	540
Db	500	TGCGAGCCAAATGGCTGTGATCAAAAATTCAACACAAATTAACCTTGAAGAACATTTTGA	559
QY	541	ACGGAACATGAAAAATCAACAAAACAAATATATATGCACTTTTGAAGCTGTAAGAGAC	600
Db	560	ACGGAACATGAAAAATCAACAAAACAAATATATATGCACTTTTGAAGCTGTAAGAGAC	619
QY	601	CTTTAAGAAACATCAGCAGCTGAAATTCGATGCAGTCCAGCATACCAATGAACTCTATT	660
Db	620	CTTTAAGAAACATCAGCAGCTGAAATTCGATGCAGTCCAGCATACCAATGAACTCTATT	679
QY	661	CAACTGTACCCAGAGAGATGTGGGAACACTTTGCATCACCACGAAGCTGAAACGACA	720
Db	680	CAACTGTACCCAGAGAGATGTGGGAACACTTTGCATCACCACGAAGCTGAAACGACA	739
QY	721	TGCCAAGGCCACAGAGGCCATATGATGTCAAAAAGGATSTGCTTTGGGCAAAAACATG	780
Db	740	TGCCAAGGCCACAGAGGCCATATGATGTCAAAAAGGATSTGCTTTGGGCAAAAACATG	799
QY	781	GACGGAACCTTCTGAACATGTGAGAGAAACCCATAAAGGAAATATCATGTGAAGTATG	840
Db	800	GACGGAACCTTCTGAACATGTGAGAGAAACCCATAAAGGAAATATCATGTGAAGTATG	859
QY	841	CCGGAACCACTTTAAGCGCAAGATTACCTTAAAGCAACATGAAGAACTCATGCCCCAGA	900
Db	860	CCGGAACCACTTTAAGCGCAAGATTACCTTAAAGCAACATGAAGAACTCATGCCCCAGA	919
QY	901	AAGGATGTATGTGCGCTGTCCAGAGAAAGGCTGTGGAACAACCTATACCTGTTTAA	960
Db	920	AAGGATGTATGTGCGCTGTCCAGAGAAAGGCTGTGGAACAACCTATACCTGTTTAA	979
QY	961	TCTCCAAAGCCATATCCTCTCTCTTCATAGGAAAGCCGCTTTTGTGTGAACATGC	1020
Db	980	TCTCCAAAGCCATATCCTCTCTCTTCATAGGAAAGCCGCTTTTGTGTGAACATGC	1039
QY	1021	TGCGTGTGGCAAAACATTTGCAATGGAACAAAGTCTCACTAGGATGCTGTTGTACATGA	1080
Db	1040	TGCGTGTGGCAAAACATTTGCAATGGAACAAAGTCTCACTAGGATGCTGTTGTACATGA	1099
QY	1081	TTCGCAAGCAAGAAATGAAGCTCAAAAGTCAAAAAATTCCTGTGAANAACGAGATTGGC	1140
Db	1100	TTCGCAAGCAAGAAATGAAGCTCAAAAGTCAAAAAATTCCTGTGTGAANAACGAGATTGGC	1159
QY	1141	CTCTCATCTCAGTGAATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGAATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTGTG	1219
QY	1201	TCAAAACGAGAGTCAACCACTGTGTGGAAGCAAGATGCTCTCGACAGTTGACATCT	1260
Db	1220	TCAAAACGAGAGTCAACCACTGTGTGGAAGCAAGATGCTCTCGACAGTTGACATCT	1279
QY	1261	TACGCTTGAGCTAA 1273	
Db	1280	TACGCTTGAGCTAA 1292	

```

RESULT 2
US-09-724-676A-44689
: Sequence 44689, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724.676A

```

```

:
: CURRENT FILING DATE: 2000-11-28
:
: NUMBER OF SEQ ID NOS: 97222
:
: SOFTWARE: patentIn version 3.2
:
: SEQ ID NO 44689
:
: LENGTH: 1906
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: misc.feature
:
: LOCATION: (1814)..(1814)
:
: OTHER INFORMATION: n is a,c,g, or t
:
: US-09-724-676A-44689

```

Query Match	99.6%;	Score 1268.2;	DB 5;	Length 1906;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1270;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0;

QY	1	ATGCGCAGACGAGCGCGCGACGCGGGGGGTGCTCTGTGATACCGCGCGCTCCGGAAGT	60
Dp	20	ATGCGCAGACGAGCGCGCGCGACGCGGGGGGTGCTCTGTGATACCGCGCGCTCCGGAAGT	79
QY	61	GTGCGGCGCTGCGCGGAAAGTTTCAGCAGGAGACGCTGGGCGCGGCGCGGTTCCGGC	120
Dp	80	GTGCGGCGCGCGCGCGGAAAGTTTCAGCAGGAGACGCGTGGGCGCGGCGCGGTTCCGGC	139
QY	121	ACGTGTCTTCGGACGTGGCGACGCGCCCTTGCCCTGGGCTTGGAGGCGCGCGGCGCTTGA	180
Dp	140	ACGTGTCTTCGGACGTGGCGACGCGCCCTTGCCCTGGGCTTGGAGGCGCGCGGCGCTTGA	199
QY	181	TTCGCGCGCGCGGTGCTGCGCGAGTCCGGTGTCTGCTTTGACATTCGCGCGAGCGCTTCATTGC	240
Dp	200	TTCGCGCGCGCGGTGCTGCGCGAGTCCGGTGTCTGCTTTGACATTCGCGCGAGCGCTTCATTGC	259
QY	241	AGCGCGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCTTCCGAGAGGTTCACTCG	300
Dp	260	AGCGCGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCTTCCGAGAGGTTCACTCG	319
QY	301	CTCCCTTCCCTGACTCAGAGGCGCAATTACAGCAAGCGTTGGAAGCTTGAAGCGCACCTGGG	360
Dp	320	CTCCCTTCCCTGACTCAGAGGCGCAATTACAGCAAGCGTTGGAAGCTTGAAGCGCACCTGGG	379
QY	361	CAAGCACAAGGGGGGAGAGACCATTTTGTGTGACTATGAAGGGGTGTGGCAAGGCTTCAT	420
Dp	380	CAAGCACAAGGGGGGAGAGACCATTTTGTGTGACTATGAAGGGGTGTGGCAAGGCTTCAT	439
QY	421	CAGGACATCACTCTGAGACCGCGCAATTTCTGACTCAACAGAGAAAGACCGTTGTGTTGG	480
Dp	440	CAGGACATCACTCTGAGACCGCGCAATTTCTGACTCAACAGAGAAAGACCGTTGTGTTGG	499
QY	481	TGCACGCCACTGGCTGTGATCAAAAAATTCACAACCAAAATCAAACTTGAAGAAACATTTTGA	540
Dp	500	TGCACGCCACTGGCTGTGATCAAAAAATTCACAACCAAAATCAAACTTGAAGAAACATTTTGA	559
QY	541	ACGCAAAACATGAAATCAACAAAAACAAATATATATGACGTTTTTGAAGACTGTAAAGAAC	600
Dp	560	ACGCAAAACATGAAATCAACAAAAACAAATATATATGACGTTTTTGAAGACTGTAAAGAAC	619
QY	601	CTTTAAGAAACATCGACGCTGAAATATCCATAGTGGCCGCAATACCAGTAAGACCTCTATT	660
Dp	620	CTTTAAGAAACATCGACGCTGAAATATCCATAGTGGCCGCAATACCAGTAAGACCTCTATT	679
QY	661	CAAGTGTACCCAGGAAGGATGTGGGAACACTTTCATATACCAGCAAGCTGAAGACGACA	720
Dp	680	CAAGTGTACCCAGGAAGGATGTGGGAACACTTTCATATACCAGCAAGCTGAAGACGACA	739
QY	721	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGAGATGTTCTTTGTGCGCAAAAACATG	780
Dp	740	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGAGATGTTCTTTGTGCGCAAAAACATG	799
QY	781	GACGGAACTTCTGAAACATGTGTACAGAAACCCATAAAGGAAATATCTATGTGAAGTATG	840
Dp	800	GACGGAACTTCTGAAACATGTGTAGAGAAACCCATAAAGGAAATATCTATGTGAAGTATG	859

QY	841	CCGAAACGATTTAAAGCGAAAGATTACCTTAAAGCAACACATGAAACAACTCATGCCCCAGA	900
Db	860	CCGGAAACATTTTAAAGCGCAAGATTACCTTAAAGCAACACATGAAACAACTCATGCCCCAGA	919
QY	901	AAGGATGTATGTGCGTGTCCAGAGAGAGCGCTGTGAGAGAACCTATACACTGTGTGTTAA	960
Db	920	AAGGATGTATGTGCGTGTCCAGAGAGAGCGCTGTGAGAGAACCTATACAACTGTGTGTTAA	979
QY	961	TCTCCAAAGGCATATTCCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTGAACATGC	1020
Db	980	TCTCCAAAGGCATATTCCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGTGAACATGC	1039
QY	1021	TGGCTGTGGCAAAACATTGGCAATGAGAAACAAGCTCCTACAGCATGCTGTGTATCATGA	1080
Db	1040	TGGCTGTGGCAAAACATTGGCAATGAGAAACAAGCTCCTACAGCATGCTGTGTATCATGA	1099
QY	1081	TCCTGCACAAAGAAATGAAGCTCAAAAGTCAAAAAAATCTGTGAAAAACGGAAGTTGGC	1140
Db	1100	TCCTGCACAAAGAAATGAAGCTCAAAAGTCAAAAAAATCTGTGAAAAACGGAAGTTGGC	1159
QY	1141	CTCTCATCTCAGTGGATATATATCCCTCCCAAGAGAAACAAAGGCAAGGCTTATCTTTGG	1200
Db	1160	CTCTCATCTCAGTGGATATATATCCCTCCCAAGAGAAACAAAGGCAAGGCTTATCTTTGG	1219
QY	1201	TCAAAACGAGAGTCACCCAACTGTGTGAGAAACAAGATGCTTCGACAGATTGCACTACT	1260
Db	1220	TCAAAACGAGAGTCACCCAACTGTGTGAGAAACAAGATGCTTCGACAGATTGCACTACT	1279
QY	1261	TACCTTGGCTTAA 1273	
Db	1280	TACCTTGGCTTAA 1292	

```

: RESULT 3 34-676-44634
: US-09-724-676-44634
: Sequence 44634, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181,4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44634
: LENGTH: 2019
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: (1927)..(1927)
: OTHER INFORMATION: n is a,c,g, or t
: US-09-724-676-44634

```

Query Match	99.6%	Score 1268.2	DB 5	Length 2019
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1270	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

QY	1	ATGCGCAGCAGCGGCGCCGAGCGCGGGCGGTGCTGTGACCGCGCGCGCTCCCGGAACT	60
Db	20	ATGGCCAGCAGCAGCGCGCGCGCGAGCGCGGGCGGGTGCCTGGTACCGCGCGCGTCCCGGAACT	79
QY	61	GTGCGCGGTGCGCGCGCGAAAGGTACAGAGGGAAGCCGTGTGGCGCGGGCGCGCGGTTCCTGGGC	120
Db	80	GTGCGCGGGCGCGCGCGCGAAAGTTACAGCGGGAAGCCGTGGCGCGGGCGCGCGGTTCCTGGGC	139
QY	121	ACGTGTCTCGGACAGTGGGACAGCGCGGCTTGCGCCCTGGAGGCTGGAAGCGCGCGCCCTTGA	180
Db	140	ACGTGTCTCGGACAGTGGGACAGCGCGGCTTGCGCCCTGGAGGCGCGCGCCCTTGA	199
QY	181	TTCGCGCGGCGGTGGTCCGCCAGTGGGTGTGTGCTCTTACCATGCGGAGCGGTTCAATTGC	240
Db	200	TTCGCGCGGCGGTGGTCCGCCAGTGGGTGTGTGCTCTTACCATGCGGAGCGGTTCAATTGC	259

OY	241	AGCCGGAGAGAGCTCACCTCCGACCCCGCGCCCGCGCTTCCAGAGAGTTTCATCTG	300
Db	260	AGCCGGGAGAGAGCTCACAGTCCGACCCCGCGCCCGCGCTTCCAGAGAGTTTCATCTG	319
OY	301	CTCTCTCCCTACTGTCAGCGCCCAATTACAGACAAAGCCTTGAAGCTTGAAGCGCACCTGTG	360
Db	320	CTCTCTCCCTACTGTCAGCGCCCAATTACAGCAAAAGCCTTGAAGCTTGAAGCGCACCTGTG	379
OY	361	CAACACACACGGGGAGAGACACATTTTGTGTGACTATGAAAGGTGTGGCAAGGCTTCAT	420
Db	380	CAACACACACGGGGAGAGACACATTTTGTGTGACTATGAAAGGTGTGGCAAGGCTTCAT	439
OY	421	CAGGAGCTACCATGTGAGCCGCCACATTTGCACTACACAGAGAGAAAGCCGTTGTTTG	480
Db	440	CAGGAGCTACCATGTGAGCCGCCACATTTGCACTACACAGAGAGAAAGCCGTTGTTTG	499
OY	481	TGCAAGCACTGAGCTGTGATCAAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGA	540
Db	500	TGCAAGCAATGAAATTTGATCAAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGA	559
OY	541	ACGCAACATGAAATTTGATCAAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGA	600
Db	560	ACGCAACATGAAATTTGATCAAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGA	619
OY	601	CTTTAAGAAACATACAGAGCTGAAAAATTCGATCAGTCCAGCATACCAATGAAACCTCTATT	660
Db	620	CTTTAAGAAACATACAGAGCTGAAAAATTCGATCAGTCCAGCATACCAATGAAACCTCTATT	679
OY	661	CAAGTGTACCCAGAGAGATGTGGGAAACACTTTGCATCACCAGCAGAGCTGAAGACACA	720
Db	680	CAAGTGTACCCAGAGAGATGTGGGAAACACTTTGCATCACCAGCAGAGCTGAAGACACA	739
OY	721	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGCATCTTCTTGTGGCAAAAACATG	780
Db	740	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGCATCTTCTTGTGGCAAAAACATG	799
OY	781	GACGGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG	840
Db	800	GACGGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG	859
OY	841	CCGGAAAAATTTAAGGCAAAAGTTTACCTTGAAGACATGAAACATCATGCCCCCGACA	900
Db	860	CCGGAAAAATTTAAGGCAAAAGTTTACCTTGAAGACATGAAACATCATGCCCCCGACA	919
OY	901	AAGGATGTATGTGCTGTCCAGAGAGAGCTGTGAGAAAGCTTACTACTGTGTTTAA	960
Db	920	AAGGATGTATGTGCTGTCCAGAGAGAGCTGTGAGAAAGCTTACTACTGTGTTTAA	979
OY	961	TCTCCAAAGCATATCTCTTCCTTCCATGAGAGAAAGCCCTTTTGTGTGAACATGC	1020
Db	980	TCTCCAAAGCATATCTCTTCCTTCCATGAGAGAAAGCCCTTTTGTGTGAACATGC	1039
OY	1021	TGGCGTGGCAAAACATTTGCATGAAACAAATCTCATTAGGCAAGCGTTGTGCATGA	1080
Db	1040	TGGCGTGGCAAAACATTTGCATGAAACAAATCTCATTAGGCAAGCGTTGTGCATGA	1099
OY	1081	TCTTGACAGAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGAAGAAACGGAGTTTGGC	1140
Db	1100	TCTTGACAGAGAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGAAGAAACGGAGTTTGGC	1159
OY	1141	CTCTCATCTCTGAGATATATATCCCTCCCAAAAGAGAAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCTGAGATATATATCCCTCCCAAAAGAGAAAGGCAAGGCTTATCTTTGTG	1219
OY	1201	TCAAAAGGAGAGTCAACCAACTGTGTGAAGACAGAGTCTCTGACAGTTGCAGTACT	1260
Db	1220	TCAAAAGGAGAGTCAACCAACTGTGTGAAGACAGAGTCTCTGACAGTTGCAGTACT	1279
OY	1261	TACCTTGTGCTTAA 1273	
Db	1280	TACCTTGTGCTTAA 1292	

```

RESULT 4
US-09-724-676A-44634
: Sequence 44634, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44634
: LENGTH: 2019
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1927)..(1927)
: OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-44634

```

Query Match	99.68;	Score 1268.2;	DB 5;	Length 2019;
Best Local Similarity	99.88;	Pred. No. 0;		
Matches 1270; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

OY	1	ATGGCAGCAGCAGCGGCGGAGCGCGGGGCGGTGCTGTGTGACCGCGCGCGCTCCCGGAAGT	60
Db	20	ATGGCGACAGCAGCGGCGCGGAGCGCGGGGCGGTGCTGTGTGACCGCGCGCGCTCCCGGAAGT	79
OY	61	GTGCCGCGCGGTGCGCGGAAGTTCAAGCAGGAGCGCTGTGGCGCGGGCGCGCGTTCGCGGC	120
Db	80	GTGCCGCGCGCGCGCGGAAGTTCAAGCAGGAGCGCTGTGGCGCGGGCGCGCGTTCGCGGC	139
OY	121	ACGTGTCTCGGCGAGGTGACACGCGCGCTGGGCGCTTGAGAGGGCGCGGCGCGCTGGA	180
Db	140	ACGTGTCTCGGCGAGGTGACACGCGCGCTGGGCGCTTGAGAGGGCGCGGCGCGCTGGA	199
OY	181	TCGCGCGCGCGGTGTGCGCGAGTCCGTGTCTCCCTTGACCATCGCGCGCGCTTCATTTGC	240
Db	200	TCGCGCGCGCGGTGTGCGCGAGTCCGTGTCTCCCTTGACCATCGCGCGCGCTTCATTTGC	259
OY	241	AGCGGCGGAGAGCTCAGTCTCGACCGCGCGCGCGCGCGCGCTTCCGAGAGGTTCAATCTG	300
Db	260	AGCGGCGGAGAGCTCAGTCTCGACCGCGCGCGCGCGCGCGCTTCCGAGAGGTTCAATCTG	319
OY	301	CTCCCTTCCCTGACTGCGAGCGCCATTACAGCAAAAGCCTTGAGAGCTTGACCGACCTGTG	360
Db	320	CTCCCTTCCCTGACTGCGAGCGCCATTACAGCAAAAGCCTTGAGAGCTTGACCGACCTGTG	379
OY	361	CAGGACACAGGGGGAGAGACCATTTGTTGTGACTATGAAGGGGTGTGGCAAGGCTTCAT	420
Db	380	CAGGACACAGGGGGAGAGACCATTTGTTGTGACTATGAAGGGGTGTGGCAAGGCTTCAT	439
OY	421	CAGGAGTACCATCTGAGCCGCCACATTTGTGACTCACAGAGAGAAAGCGTTTGTTTG	480
Db	440	CAGGAGTACCATCTGAGCCGCCACATTTGTGACTCACAGAGAGAAAGCGTTTGTTTG	499
OY	481	TGCAGCCACTGGCTGTGTATCAAAAATTCACACAAATTCAAACTTGAGAGAACATTTTGA	540
Db	500	TGCAGCCCAATGGCTGTGTATCAAAAATTCACACAAATTCAAACTTGAGAGAACATTTTGA	559
OY	541	ACGCAAAATCATGAAATCAACAAAAACATATATGTGAGTTTGAACAGCTGTAAAGAGAC	600
Db	560	ACGCAAAATCATGAAATCAACAAAAACATATATGTGAGTTTGAACAGCTGTAAAGAGAC	619
OY	601	CTTTTACGAACATCAGCAGCTGAAATTCATCAGTGCACCATTACCAATGAACCTCTATT	660
Db	620	CTTTTACGAACATCAGCAGCTGAAATTCATCAGTGCACCATTACCAATGAACCTCTATT	679
OY	661	CAAGGTGCCGAGGAAGATGTGGCAAAACCTTGCATCCACGACGAAGCTGAAAGACGACA	720
Db	680	CAAGGTGCCGAGGAAGATGTGGCAAAACCTTGCATCCACGACGAAGCTGAAAGACGACA	739

QY	721	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGGATGTCCTTTGGGCAAAAAACATC	780
Db	740	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGGATGTCCTTTGGGCAAAAAACATC	799
QY	781	GACCGAACTTCTGAACATGTGAGCAAAACCCATAAAGAGAAATCTATGTGAAGTATG	840
Db	800	GAGCGAACTTCTGAACATGTGAGCAAAACCCATAAAGAGAAATCTATGTGAAGTATG	859
QY	841	CCGGAATAACATTTAAACGCAAGATTACCTTAAGCAACATGAATACTATGCCCCAGA	900
Db	860	CCGGAATAACATTTAAACGCAAGATTACCTTAAGCAACATGAATACTATGCCCCAGA	919
QY	901	AAGGATGTATGTGCGCTGTCCAGACAAGCGTGTGAANAACCTTACTACTGTGTTAA	960
Db	920	AAGGATGTATGTGCGCTGTCCAGACAAGCGTGTGAANAACCTTACTACTGTGTTAA	979
QY	961	TCTCTCAAGCATATCCCTCTCTCTTCATGAGAAAGCCGCCCTTTGTGTGTGAACATGC	1020
Db	980	TCTCTCAAGCATATCCCTCTCTCTTCATGAGAAAGCCGCCCTTTGTGTGTGAACATGC	1039
QY	1021	TGCGCTGTGGCAAAACATTTTGCATGAAACAAAGTCTACCTAGCATGCTGTTGTATCATGA	1080
Db	1040	TGCGCTGTGGCAAAACATTTTGCATGAAACAAAGTCTACCTAGCATGCTGTTGTATCATGA	1099
QY	1081	TCTCTGAAAGAAAGAAAATGAAGTCAAAATGCAAAAAATTCGTGTGAAAAAAGGAGTTGGC	1140
Db	1100	TCTCTGAAAGAAAGAAAATGAAGTCAAAATGCAAAAAATTCGTGTGAAAAAAGGAGTTGGC	1159
QY	1141	CTCTCATCTCAGTGGATATATATCCCTCCCAAAAAAGGAACAAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGGATATATATCCCTCCCAAAAAAGGAACAAAGGCAAGGCTTATCTTTGTG	1219
QY	1201	TCAAAAGGAGAGTCAACCAACTGTGTGGAAGACAAGATGCTCTCGACAGTTGACAGTACT	1260
Db	1220	TCAAAAGGAGAGTCAACCAACTGTGTGGAAGACAAGATGCTCTCGACAGTTGACAGTACT	1279
QY	1261	TACCTTTGGCTAA 1273	
Db	1280	TACCTTTGGCTAA 1292	

RESULT 5
US-09-724-676-44645
: Sequence 44645, Application US/09724676

```

: APPLICANT: CompuGen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44645
: LENGTH: 2104
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-724-676-44645

```

Query Match	99.6%	Score 1266.2	DB 5	Length 2104
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1270	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

QY 1 ATGGCGACAGCGGCGCGCGAGCGCGGGCGGTCCTGTGTGACCGCGCGCGCTCCCGAGT 60
Db 20 ATGGCGACACCGGCGCGCGAGCGCGGGCGGTGCTGTGTGACCGCGCGCGCTCCCGAGT 79	
QY 61 GTGGCGGCGTGGCGCGGAAGTTACACAGGAGCGCGTGGCGCGCGCGCGCGTTCGCCGC 120	
Db 80 GTGGCGGCGCGCGCGAGGTTACACAGGAGCGCGTGGCGCGCGCGCGCGGTTCCCGC 139	
QY 121 ACGTGTCTGGCGACGTGGACGCGCGCTGGGCTCTGGAGGCGCGCGCGCTCGA 180	

Db	140	ACGTCGCGACAGTCGACAGCGCGCTGGCCCTGGGCTTGGAGGCGCGCGCCCTCGGA	199
Qy	181	TCGCACCGGCGTGTCGCACGTCGCGTGTCTGTCTGACATCGCGCAGCGCTTCATTC	240
Db	200	TCGCGCGCGCGTGTCGCGACGTCCGATGTCTGTCTTGACATCGCGCAGCGCTTCATTC	259
Qy	241	AGCCGCGAGAGCTCACTCCGACCCCGCGCGCCCGCGCTTCCAGAGAGTTCATCTG	300
Db	260	AGCCGCGAGAGCTCACTCCGACCCCGCGCGCGCCCGCGCTTCCAGAGAGTTCATCTG	319
Qy	301	CTCTCTCCCTGATGTCAGCGCCCAATTACAGCAAAAGCTGTGAAGCTTGACGGCACCTGTG	360
Db	320	CTCTCTCCCTGATGTCAGCGCCCAATTACAGCAAAAGCTGTGAAGCTTGACGGCACCTGTG	379
Qy	361	CAACACACAGGGGGAGAGACATTTCTTTGTGACTGTGAAGGCTGTGGCAAGCCCTCAT	420
Db	380	CAACACACAGGGGGAGAGACATTTCTTTGTGACTGTGAAGGCTGTGGCAAGCCCTCAT	439
Qy	421	CAGGACTACCATCTGAGCGCGCACATTTCTGATCACACAGAGAAAAACCGTTGTGTG	480
Db	440	CAGGACTACCATCTGAGCGCGCACATTTCTGATCACACAGAGAAAAACCGTTGTGTG	499
Qy	481	TGCGAGCCACTGGCGTGTATATAAAATTTCAACACAAAATTAACCTTGAAGAAACATTTTGA	540
Db	500	TGCGAGCCAAATGGCTGTATATAAAATTTCAACACAAAATTAACCTTGAAGAAACATTTTGA	559
Qy	541	ACGAAACATGAAAAATACAAAAACAAATATATGCACTTTTGAAGCTGTGAAGAAAC	600
Db	560	ACGAAACATGAAAAATACAAAAACAAATATATGCACTTTTGAAGCTGTGAAGAAAC	619
Qy	601	CTTTAAGAAACATCAGCAGCTGAAATTCATCAGTCCAGCATCAATGAAACCTCATTT	660
Db	620	CTTTAAGAAACATCAGCAGCTGAAATTCATCAGTCCAGCATCAATGAAACCTCATTT	679
Qy	661	CAATGTATCCCAAGGAAGATGTGGGAAACACTTTGCATACCACAGCTGAAGACACA	720
Db	680	CAATGTATCCCAAGGAAGATGTGGGAAACACTTTGCATACCACAGCTGAAGACACA	739
Qy	721	TGCCAAGGCCACAGAGCGCTATGTATGTCAAAAAGATTTCTTTGTGGCAAAAACATG	780
Db	740	TGCCAAGGCCACAGAGCGCTATGTATGTCAAAAAGATTTCTTTGTGGCAAAAACATG	799
Qy	781	GACGGAACCTTCTGAACATGTGAGAGAAACCCATAAGAGAAATACTATGTGAAGTATG	840
Db	800	GACGGAACCTTCTGAACATGTGAGAGAAACCCATAAGAGAAATACTATGTGAAGTATG	859
Qy	841	CCGGAACCATTTTAACGCCAAAGATTAACCTTAAGCAACACATGAAGAACTATGCCCGACA	900
Db	860	CCGGAACCATTTTAACGCCAAAGATTAACCTTAAGCAACACATGAAGAACTATGCCCGACA	919
Qy	901	AAGGATGTATGTCCGCTGTCAAGAGAAAGCTGTGAACAACTTACTACTGTGTATA	960
Db	920	AAGGATGTATGTCCGCTGTCAAGAGAAAGCTGTGAACAACTTACTACTGTGTATA	979
Qy	961	TCTCCAAAGCCATTCCTCTCTCTTCATGAGAAAGCCCGCTTTGTGTGAACATATC	1020
Db	980	TCTCCAAAGCCATTCCTCTCTTCATGAGAAAGCCCGCTTTGTGTGAACATATC	1039
Qy	1021	TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCACTAGGCATGCTGTTGTACATGA	1080
Db	1040	TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCACTAGGCATGCTGTTGTACATGA	1099
Qy	1081	TCCTGACAAAGAAATGAAAGCTCAAAAGTCAAAATAATCTCTGAAAAACGAGTTTGGC	1140
Db	1100	TCCTGACAAAGAAATGAAAGCTCAAAAGTCAAAATAATCTCTGAAAAACGAGTTTGGC	1159
Qy	1141	CTCTCATCTCAGTGTATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGTATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTGTG	1219
Qy	1201	TCAAAACGAGAGTCAACCCCACTGTGTGAAGCAAGATGCTCTGACAGTTGCAGTACT	1260
Db	1220	TCAAAACGAGAGTCAACCCCACTGTGTGAAGCAAGATGCTCTGACAGTTGCAGTACT	1279

QY	1261	TACCTTGCTAA	1273	
Db	1280	TACCTTGCTAA	1292	
RESULT 6				
US-09-724-676A-44645				
; Sequence 44645, Application US/09724676A				
; GENERAL INFORMATION:				
; APPLICANT: CompuGen LTD				
; TITLE OF INVENTION: Variants of alternative splicing				
; FILE REFERENCE: 129181.4 CompuGen				
; CURRENT APPLICATION NUMBER: US/09/724.676A				
; NUMBER OF SEQ ID NOS: 97222				
; SOFTWARE: Patentin version 3.2				
; SEQ ID NO: 44645				
; LENGTH: 2104				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
US-09-724-676A-44645				
Query Match				
Best Local Similarity 99.6%; Score 1268.2; DB 5; Length 2104;				
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
QY	1	ATGCGAGCAGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGACCGCGCGCGCTCCGGAAGT	60	
Db	20	ATGCGCAGCAGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGACCGCGCGCGCTCCGGAAGT	79	
QY	61	GTGCGCGCGCTCCCGGGAAGTTCAAGAGGACCGTGGGCGCGCGCGCTTCCGGC	120	
Db	80	GTGCTTCCGGC	139	
QY	121	ACGTGCTGTGGACGTCGCGAGCGCGCGCGCGCGCGCGCGCGCTTGGAGCGCGCGCGCTTGA	180	
Db	140	ACGTGCTGTGGACGTCGCGAGCGCGCGCGCGCGCGCGCGCGCTTGGAGCGCGCGCGCTTGA	199	
QY	181	TCCGCGCGCGCGTGTGCGCGCGCGCGAGTGTGCTGCTTGAACCATCGCGCGCTTATTCG	240	
Db	200	TCCGCGCGCGCGTGTGCGCGCGCGCGAGTGTGCTGCTTGAACCATCGCGCGCTTATTCG	259	
QY	241	ACCGCGCGAGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCCAGAGGTTTCACTCG	300	
Db	260	ACCGCGCGAGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCCAGAGGTTTCACTCG	319	
QY	301	CTCTCTTCCCTGACTGTGAGCGCCCAATTACAGCAAAAGCTGGAAGCTTGAAGCGCCACTGTG	360	
Db	320	CTCTCTTCCCTGACTGTGAGCGCCCAATTACAGCAAAAGCTTGAAGCGCCACTGTG	379	
QY	361	CAGACACAGCGGGGAGAGACCAATTTGTGTGACTATGAAGGCTGTGCAAGGCTTCAT	420	
Db	380	CAGACACAGCGGGGAGAGACCAATTTGTGTGACTATGAAGGCTGTGCAAGGCTTCAT	439	
QY	421	CAGGAGCTACCATTCAGCGCGCCACATTTGTGACTCACAAGAGAAAGCCGTTGTTTG	480	
Db	440	CAGGAGCTACCATTCAGCGCGCCACATTTGTGACTCACAAGAGAAAGCCGTTGTTTG	499	
QY	481	TTCAGCAGCTGCTGTGATCAAAATTTCAACACAAATATCAAACTTGAAGAAACATTTTGA	540	
Db	500	TTCAGCAGCTGCTGTGATCAAAATTTCAACACAAATATCAAACTTGAAGAAACATTTTGA	559	
QY	541	ACGCAAACTGAAATCAACAAAAACCAATATATATGACGTTTGAAGACTGTGAAGAGC	600	
Db	560	ACGCAAACTGAAATCAACAAAAACCAATATATATGACGTTTGAAGACTGTGAAGAGC	619	
QY	601	CTTTAAGAAACATCGACACTGAAAAATCCATAGTCCGACATACCAATGAACCTCTATT	660	
Db	620	CTTTAAGAAACATCGACACTGAAAAATCCATAGTCCGACATACCAATGAACCTCTATT	679	
QY	661	CAAGTGTACCCAGGAAGATGTGGAAACATTTGCATACCCAGCAACTGAACGACA	720	

Db	680	CAATGTCATCCAGAGAAAGATGTGGAAACACTTTGCAATCCACCAAGCAAGCTGAAACGACA	739
Oy	721	TGCGAAGGCCCCACGAGGGCTATGTATGTCAAAAAGAGATGTTCTTTGTGGCAAAAACATG	780
Db	740	TGCCAAGGCCCCACGAGGGCTATGTATGTCAAAAAGAGATGTTCTTTGTGGCAAAAACATG	799
Oy	781	GAGCGAATCTTGAAACATGTGAGACAAACCCATAAAGAGAAATACATATGTGAAGTATG	840
Db	800	GACCGAATCTTGAAACATGTGAGAAACCCATAAAGAGAAATACATATGTGAAGTATG	859
Oy	841	CCGGAACAACTTTAAACGCAAAAGATTACCTTAAGCAACATGAAACAACTCATGCCCCAGA	900
Db	860	CCGGAACAACTTTAAACGCAAAAGATTACCTTAAGCAACATGAAACAACTCATGCCCCAGA	919
Oy	901	AAGCAATGTATGTGCGCTGTCCAAAGAGAAGGCTGTGAGAAACCTATWACTACTGTGTTTAA	960
Db	920	AAGCAATGTATGTGCGCTGTCCAAAGAGAAGGCTGTGAGAAACCTATWACTACTGTGTTTAA	979
Oy	961	TCTCCAAAGCCATATCTCTCTTCATTCATGAGAAAGCCCCCTTTGTGTGTGAAACATGC	1020
Db	980	TCTCCAAAGCCATATCTCTCTTCATTCATGAGAAAGCCCCCTTTGTGTGTGAAACATGC	1039
Oy	1021	TGGCGTGTGCAAAAACATTTTCAATGAAACAAAGCTCTCACTAGCAWGCCTGTGTACATGA	1080
Db	1040	TGGCGTGTGCAAAAACATTTTCAATGAAACAAAGCTCTCACTAGCAWGCCTGTGTGTACATGA	1099
Oy	1081	TCCTGCACAGAGAAAATGAAAGCTCAAAAGTCAAAAAATCTCGTGAAAAAACGAGCTTGGC	1140
Db	1100	TCCTGCACAGAGAAAATGAAAGCTCAAAAGTCAAAAAATCTCGTGAAAAAACGAGCTTGGC	1159
Oy	1141	CTCTCATCTCAGTGGATATATGCTGCTCCAAAAGGAAACAAAGGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGGATATATGCTGCTCCAAAAGGAAACAAAGGGCAAGGCTTATCTTTGTG	1219
Oy	1201	TCAAAACGAGAGAGTCACCAACTGTGTGGAAGACAGATGCTCTCGACAGTTGCGAGTACT	1260
Db	1220	TCAAAACGAGAGAGTCACCAACTGTGTGGAAGACAGATGCTCTCGACAGTTGCGAGTACT	1279
Oy	1261	TACCTTTGGCTAA 1273	
Db	1280	TACCTTTGGCTAA 1292	

```

RESULT 7
US-09-724-676-44678
: Sequence 44678, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: CompuGen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: Patentin version 3.2
: SEQ ID NO 44678
: LENGTH: 2239
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-724-676-44678

```

Query Match	99.68;	Score 1268.2;	DB 5;	Length 2239;
Best Local Similarity	99.88;	Pred. NO. 0;		
Matches 1270; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

Qy	Db	Qy	Db	Qy	Db
1	20	61	80	121	180
ATTGGCACCACGGGCCCCGAGCGGGGGGGGGTGGTGTGACGCGCGCGCTTCCCGGAAT	ATGCGCAGCAGCGGCGCGGACCGGGGGGGTGGCTTGGTGAACGCGCGCGCTCCCGGAAAT	GTGCGCGCTGTGCGCGGAAGTTTACAGCAGGAGACCCCTGTGGCGCGCGCGCGTTCGCCGAC	GTGCGCGCGCGCGCGGAAGTTTACAGCAGGAGACCCCTGTGGCGCGCGCGCGTTCGCCGAC	ACGTGTCTCTGGCAGCTGTGACACGCGCGCGCTGTGGGCTCTTGGAGAGGCGCGCGCGCTCGA	ATGCGCAGCAGCGGCCCCGAGCGGGGGGGGGTGGTGTGACGCGCGCGCTTCCCGGAAT

D	b	140	ACGTCCTCGGCACGTGGTGCACGCGCGCCCTGGCCCTTGGAGGCGCGCGCCCTCGGA	199
O	y	181	TCGCGCGGCGCTGGTCTCCCGCAGTCGCGTGTCTCTTGACATTCGCGCAGCGCTTCATTTGC	240
D	b	200	TCGCGCGGCGCTGGTCCGCGAGTCGGTGTCTCTTGACATTCGCGCAGCGCTTCATTTGC	259
O	y	241	AGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCGAGAGAGGTTTCATCTG	300
D	b	260	AGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCGAGAGAGGTTTCATCTG	319
O	y	301	CTCCTTCCTTACCTGTCAGCGCCCAATTACGCAAAAGCCTTGAAGCTTGACGGCACCTGTG	360
D	b	320	CTCCTTCCTTACCTGTCAGCGCCCAATTACGCAAAAGCCTTGAAGCTTGACGGCACCTGTG	379
O	y	361	CAAGCACACGGGGAGAGACATTTGTTGTGCTATGAAAGGTGTGGCAAGCGCTTCAT	420
D	b	380	CAAGCACACGGGGAGAGACATTTGTTGTGCTATGAAAGGTGTGGCAAGCGCTTCAT	439
O	y	421	CAGGACTACCATCTGAGCGCGCACATTCGTGACTCAGCAGGAGAAAAGCCGTTGTTTG	480
D	b	440	CAGGACTACCATCTGAGCGCGCACATTCGTGACTCAGCAGGAGAAAAGCCGTTGTTTG	499
O	y	481	TGCAAGCCACTGGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAAGAAACATTTTGA	540
D	b	500	TGCAAGCCAAATGGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAAGAAACATTTTGA	559
O	y	541	ACGCAAACTGAATAATCAAAAAACAAATTAATGCACTTTTGAAGAGCTGTAAACAACAC	600
D	b	560	ACGCAAACTGAATAATCAAAAAACAAATTAATGCACTTTTGAAGAGCTGTAAACAACAC	619
O	y	601	CTTTAAGAATACATAGCAGCTGAAAATTCATCAGTCCAGCATACCAATGAACCTCTATT	660
D	b	620	CTTTAAGAATACATAGCAGCTGAAAATTCATCAGTCCAGCATACCAATGAACCTCTATT	679
O	y	661	CAAGTGTACCCAGGAGAGATGTGGAAACACTTTGCTATCCACGCAAGCTGAACGACA	720
D	b	680	CAAGTGTACCCAGGAGAGATGTGGAAACACTTTGCTATCCACGCAAGCTGAACGACA	739
O	y	721	TGCGAAGGCCACGAGGCGTATGATGTCAAAAAGATGTCCTTTGGCAAAAACATG	780
D	b	740	TGCGAAGGCCACGAGGCGTATGATGTCAAAAAGATGTCCTTTGGCAAAAACATG	799
O	y	781	GACGGAACCTTCTGAAAGCATGTGAGAGAAACCCATTAAGAGAGAAATACTATGTGAAGTATG	840
D	b	800	GACGGAACCTTCTGAAAGCATGTGAGAGAAACCCATTAAGAGAGAAATACTATGTGAAGTATG	859
O	y	841	CCGGAACCATTTTAACGCAAAAGATTACCTTAAGCAACATGAAAACCTCATGCCCCAGA	900
D	b	860	CCGGAACCATTTTAACGCAAAAGATTACCTTAAGCAACATGAAAACCTCATGCCCCAGA	919
O	y	901	AAGGATGTATGTCGTCGTCCCAAGAGAGGCTGTGAAACACTTACTACTGTGTTTAA	960
D	b	920	AAGGATGTATGTCGTCGTCCCAAGAGAGGCTGTGAAACACTTACTACTGTGTTTAA	979
O	y	961	TCTCCAAAGCCATATCCCTCTCTTCCATGAGAAACCCGCTTTTGTGTGTGAACATGC	1020
D	b	980	TCTCCAAAGCCATATCCCTCTCTTCCATGAGAAACCCGCTTTTGTGTGTGAACATGC	1039
O	y	1021	TGGCTGTGGCAAAACATTTGCAATGAAACAAAGCTCTACTAGGCATGCTGTGTATCATGA	1080
D	b	1040	TGGCTGTGGCAAAACATTTGCAATGAAACAAAGCTCTACTAGGCATGCTGTGTATCATGA	1099
O	y	1081	TCCTGACAAAGAAATAATGAAGCTCAAAAGTCAAAAAATCTCGTGA AAAACGAGATTTGGC	1140
D	b	1100	TCCTGACAAAGAAATAATGAAGCTCAAAAGTCAAAAAATCTCGTGA AAAACGAGATTTGGC	1159
O	y	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGCAAAAGGCAAGGCTTATCTTTGTG	1200
D	b	1160	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGCAAAAGGCAAGGCTTATCTTTGTG	1219
O	y	1201	TCAAAACGAGAGTCAACCAACTGTGTGAAGCAAGATGCTCTCGACAGTTGCAGTACT	1260

Dd	1220	TCAAAAGGAGAGATCACCACACTGTGTGGAGACAGATGCTCTCGACAGTTGGAGTACT	1279
QY	1261	TACCCCTTGGCTAA	1273
Dd	1280	TACCTTGGCTAA	1292

```

RESULT 8
US-09-724-676A-44678
; Sequence 44678, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44678
; LENGTH: 2239
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-724-676A-44678

```

Query Match	99.6%	Score	1268.2	DB 5	Length	2239
Best Local Similarity	99.8%	Pred. No.	0			
Matches 1270; Conservative	0	Mismatches	3		Indels	0
					Gaps	0

OY	1	ATGGCAGACAGGCGCGCCGAGACGGCGGGCGGTCCTGGTGTACCGCGCGCGCTCCCGGAAGT	60
Db	20	ATGGGCACACAGCGCGCGCGAGCGCGGGCGGTGGCTGGTGTACCGCGCGCGCTCCCGGAAGT	79
OY	61	GTGCGGCGGTGCGCGCGAAGGTTTCAGCAGGGAGCCGTGGCGCGGGCGCGCGGTTCGCGAGC	120
Db	80	GTGCGCGCGCGCGCGCGAAGTTTCAGCAGGGAGCCGTGGCGCGGGCGCGCGGTTCGCGAGC	139
OY	121	ACGTGTCTCGGCGACGTGGCAGCGCGCCCTGGCCCTGGGCTTGGAGGGCGCGGGCCCTCGA	180
Db	140	ACGTGTCTCGGCGACGTGGCAGCGCGCCCTGGCCCTGGGCTTGGAGGGCGCGGGCCCTCGA	199
OY	181	TCGCGCGCGCGCTGGTTCGCGCGAGTGGGTGGTGTGCTCTTACCATCATGCGCCAGACGTTTCATTCG	240
Db	200	TCGCGCGCGCGCTGGTTCGCGCGAGTGGGTGGTGTGCTCTTACCATCATGCGCCAGACGTTTCATTCG	259
OY	241	AGCGGCGGAGAGCTACGTCCGACCCCGCGCGCCCGCGCTTCCAGAGGTTTCATCTG	300
Db	260	AGCGCGGAGAGCTACGTCCGACCCCGCGCGCCCGCGCTTCCAGAGGTTTCATCTG	319
OY	301	CTCCCTTCCCTACGTGACGAGCGGCATTTCACGAAGCCTGGAAGTTCAGCGGCACCTGTG	360
Db	320	CTCCCTTCCCTACGTGACGAGCGGCATTTCACGAAGCCTGGAAGTTCAGCGGCACCTGTG	379
OY	361	CAAGCAGCAGGGGGGAGACCACTTTGTTTGTGACTATCAAGGGGTGTGGCAAGGCGCTTCAT	420
Db	380	CAAGCAGCAGGGGGGAGAGACCATTGTTTGTGACTATGAAGGGGTGTGGCAAGGCGCTTCAT	439
OY	421	CAGGAGCTACCATCTGAGCGCCGCACATTTGTGACTACACAGAGAGAAAAGCCGTTTGTG	480
Db	440	CAGGAGCTACCATCTGAGCGCCGCACATTTGTGACTACACAGAGAGAAAAGCCGTTTGTG	499
OY	481	TGCACCCACTCGGTGTGATCAAAAATTCAACACCAAAATCAAACTTGAAGAAACATTTTGA	540
Db	500	TGCACCCCAATGGGTGTGATCAAAAATTCAACACCAAAATCAAACTTGAAGAAACATTTTGA	559
OY	541	ACGCAACATGAAAAATCAACAAAAACAATATATATGCAAGTTTGAAGAGCTGTAAAGAGAC	600
Db	560	ACGCAACATGAAAAATCAACAAAAACAATATATATGCAAGTTTGAAGAGCTGTAAAGAGAC	619
OY	601	CTTTAAGAAATCATGCGAGCTGAAAAATCATCATGTGCAGATTCGAAATCAACCTCATTT	660
Db	620	CTTTAAGAAATCATGCGAGCTGAAAAATCATCATGTGCAGATTCGCAATGAAACCTCATTT	679
OY	661	CAAGTGTACCCAGAGAGGATGTGGAAAAACCTTTGCATTCACCCAGCAAGCTGAAGACACA	720

Db	680	CAAGTGTACCCAGSAGAGATGTGGGAAACACTTTTGATGCCATGCCACCAAGCTGAACGACA	739
QY	721	TGCGAAGGCCCCAGAGGGCTATGTATGTCAAAAAGGATGTTCCTTTGTGCAAAAACATG	780
Db	740	TGCCAAGGCCCCACAGAGGCTATGTATGTCAAAAAGGATGTCTCTTTGTGCAAAAACATG	799
QY	781	GAGGCACTTCTGAAACATGTGAGAAACCCATAAAGGAAATACATGTGAAAGTATG	840
Db	800	GAGGCACTTCTGAAACATGTGAGAAACCCATAAAGGAAATACATGTGAAAGTATG	859
QY	841	CCGGAAACATTTAAACGCAAAAGATTACCTTAAGCAACATGAATACTCATGCCCCAGA	900
Db	860	CCGGAAACATTTAAACGCAAAAGATTACCTTAAGCAACATGAATACTCATGCCCCAGA	919
QY	901	AAGGATGTATGTGCTGTCTCCAAAGAGGCTGTGGAAGAACTTACTACTGTGTTAA	960
Db	920	AAGGATGTATGTGCTGTCTCCAAAGAGGCTGTGGAAGAACTTACTACTGTGTTAA	979
QY	961	TCTCCAAAGCCATATCCCTCTCTCTTCGATGAGAAACCCGCCCTTTTGTGTGTGAACATGC	1020
Db	980	TCTCCAAAGCCATATCCCTCTCTTCCTTCATGAGAAACGCCGCCCTTTTGTGTGTGAACATGC	1039
QY	1021	TGGCTGTGSCAAAACATTGTGCAATGAACCAAACTCTCATAGGCACTGCTGTTGTACATGA	1080
Db	1040	TGGCTGTGSCAAAACATTGTGCAATGAACCAAACTCTCATAGGCACTGCTGTTGTACATGA	1099
QY	1081	TCTGTGACAGAGAGAAATGAAAGCTCAAAGTCAAAAAATCTCGTGA AAAACGAGCTTTGGC	1140
Db	1100	TCTGTGACAGAGAGAAATGAAAGCTCAAAGTCAAAAAATCTCGTGA AAAACGAGCTTTGGC	1159
QY	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGAAACAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGAAACAAGGCAAGGCTTATCTTTGTG	1219
QY	1201	TCAAAACGAGAGTCAACCCAAAGTGTGGAAGCAAGATGCTCTGACAGTGTGCACTACT	1260
Db	1220	TCAAAACGAGAGTCAACCCAAAGTGTGGAAGCAAGATGCTCTGACAGTGTGCACTACT	1279
QY	1261	TACCTTGTGCTTAA 1273	
Db	1280	TACCTTGTGCTTAA 1292	

```

RESULT 9
US-09-724-676-44656
; Sequence 44656, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ. ID NOS. 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44656
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-44656

```

Query Match	99.6%	Score 1268.2	DB 5	Length 2271
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1270; Conservative	0	Mismatches 3	Indels 0	Gaps 0

Qy 1 ATGCGACACAGCGGGCGCCGACGCGGGGGGCGTGGTGACCGCGCGCGCTCCCGAGAGT 60
Qy 20 ATGCGCAGAGGGGGCGACGCGGGGGGGTGGCTGGTGACCGCGCGCGCTCCCGAGAGT 79
Db
Qy 61 GTGCGCGCGTGTGCGCGGAAGTTCA6CACAGGAGCCGTG9GGCCGGGCGCGCGGTTCGGGCG 120
Db 80 GTGCGCGCGCGCGCGGAAGTTCA6CACAGGAGCCGTG9GGCGCGCGCGCGGTTCGGGCG 139

OY	121	ACGTCCTCGGACATCTGAGAGCGCCCTGGCCCTGGGCTTGGAGAGCGCGCGCCCTCGGA	180
Db	140	ACGTGCTCGGACATGTGGCAGCGCGCCCTGGCCCTTGGAGGCGCGCGCCCTCGGA	199
OY	181	TCGCGCGGCGCTGGTCCGCCAGTCGCGTGTCTCTTGACATCTGCGCGAGCGCTTCATTGCG	240
Db	200	TCGCGCGGCGCGTGTGGCCGACGTGCGTGTCTCTTGACATCTGCGCGAGCGCTTCATTGCG	259
OY	241	AGCGGGGAGAGCTCAGCTCCGAGCCCGCGCGCGCCGCTTCCGAGGAGGTTCATCTG	300
Db	260	AGCGGGGAGAGCTCAGCTCCGAGCCCGCGCGCGCCGCTTCCGAGGAGGTTCATCTG	319
OY	301	CTCCCTCCCTACTGCGAGGCGCAATTACAGCAAGCGTGAAGCTTGAGCGCCACTCTG	360
Db	320	CTCCCTCCCTACTGCGAGGCGCAATTACAGCAAGCGTGAAGCTTGAGCGCCACTCTG	379
OY	361	CAAGACACAGGGGGAGAGACATTGTTGTGTGACTGTGAAGGGTGTGGCAGAGCCCTGAT	420
Db	380	CAAGACACAGGGGGAGAGACATTGTTGTGTGACTGTGAAGGGTGTGGCAGAGCCCTGAT	439
OY	421	CAGGACTACCATCTGAGCGCCGACATCTGACTCAGACAGAGAGAAAACCGTTTGTG	480
Db	440	CAGGACTACCATCTGAGCGCCGACATCTGACTCAGACAGAGAGAAAACCGTTTGTG	499
OY	481	TGCGAGCCACTGGCTGTGATCAAAAATTCAACACAAATTAACCTTGAGAGAACTTTTGA	540
Db	500	TGCGAGCCACTGGCTGTGATCAAAAATTCAACACAAATTAACCTTGAGAGAACTTTTGA	559
OY	541	ACGCAACATGAAAAATCAACAAAAACATATATATGCACTTTTGAAGACTGTGAGAACAC	600
Db	560	ACGCAACATGAAAAATCAACAAAAACATATATGCACTTTTGAAGACTGTGAGAACAC	619
OY	601	CTTTAAGAACATCAGCAGCTGAAAAATCATCATGTCGACAGATPACCAATGAACCTCTATT	660
Db	620	CTTTAAGAACATCAGCAGCTGAAAAATCATCATGTCGACAGATPACCAATGAACCTCTATT	679
OY	661	CAAGGTATCCCGAGAGAGATGTGGGAAACCTTGCATCACCACGCAAGCTGAAGACGA	720
Db	680	CAAGGTATCCCGAGAGAGATGTGGGAAACCTTGCATCACCACGCAAGCTGAAGACGA	739
OY	721	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGAGATTTCTCTTGTGTGGCAAAACATG	780
Db	740	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGAGATTTCTCTTGTGTGGCAAAACATG	799
OY	781	GACGGAACCTTGTGAACATGTGAGAGAAACCTTAAGAGAAATPCTATGTGAAGTATG	840
Db	800	GACGGAACCTTGTGAACATGTGAGAGAAACCTTAAGAGAAATPCTATGTGAAGTATG	859
OY	841	CGGAAAAACATTTAAGCGAAAGATTACCTTAAGCAACATGAAAAACATCAGCCCCAGA	900
Db	860	CGGAAAAACATTTAAGCGAAAGATTACCTTAAGCAACATGAAAAACATCAGCCCCAGA	919
OY	901	AAGGATGTATGTGCTGCTCAAGAGAGGCTGTGAAACACCTATACTACTGTGTTAA	960
Db	920	AAGGATGTATGTGCTGCTCAAGAGAGGCTGTGAAACACCTATACTACTGTGTTAA	979
OY	961	TCCTCCAAAGCCATATCTCTCTTCCATGAGGAAGCGCGCCCTTTGTGTGTGAACATGC	1020
Db	980	TCCTCCAAAGCCATATCTCTCTTCCATGAGGAAGCGCGCCCTTTGTGTGTGAACATGC	1039
OY	1021	TGGCTGTGGCAAAACATTTGCAATGAAGAAAGTCTCATAGGACGCTGTGTGATATGA	1080
Db	1040	TGGCTGTGGCAAAACATTTGCAATGAAGAAAGTCTCATAGGACGCTGTGTGATATGA	1099
OY	1081	TCCTGACAAAGAAATGAAGCTCAAAAGTCAAAAAATCTGCGAAAAACGAGATTGTGGC	1140
Db	1100	TCCTGACAAAGAAATGAAGCTCAAAAGTCAAAAAATCTGCGAAAAACGAGATTGTGGC	1159
OY	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGGAACAAGGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGGAACAAGGGCAAGGCTTATCTTTGTG	1219
OY	1201	TCAAAACGAGAGTCAACCCAACTGTGTGCAAGCAAGATCTCTGTGCACAGTTGCACTACT	1260

DB	1220	TCAAAACGAGAGTACCCAACTGTGTGGAGACAAAGATGCTGTGACAGTTTGACGACT	1279
QY	1261	TACCCCTTGCTMA 1273	
DB	1280	TACCCCTTGCTMA 1292	
RESULT 10			
US-09-724-676A-44656			
: Sequence 44656, Application US/09724676A			
: GENERAL INFORMATION:			
: APPLICANT: Comugen LTD			
: TITLE OF INVENTION: Variants of alternative splicing			
: FILE REFERENCE: 129181.4 Comugen			
: CURRENT APPLICATION NUMBER: US/09/724.676A			
: CURRENT FILING DATE: 2000-11-28			
: NUMBER OF SEQ ID NOS: 97222			
: SOFTWARE: PatentIn version 3.2			
: SEQ ID NO 44656			
: LENGTH: 2271			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
US-09-724-676A-44656			
Query Match 99.6%, Score 1268.2; DB 5; Length 2271;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1	ATGGCCACAGCGGGCGCCGACCGGGGGCGTGCCTGTGTACCCGCCGCTCCCGGAAGT	60
DB	20	ATGCCACACAGCGGGCGCCGACCGGGGGCGTGCCTGTGTACCGCCGCTCCCGGAAGT	79
QY	61	GTGCGCGGCGTCCGCGGAAGTTTCACAGGAGCGGTGGGCGCGCGCGGTTCCCGGC	120
DB	80	GTGCGCGGCGCGCGGAAGTTTCACAGGAGCGGTGGGCGCGCGCGGTTCCCGGC	139
QY	121	ACGTGCTTCGCGACGTGGCGAGCGCCCTTGCCCTTGAGGCGCGCGCCCTTGA	180
DB	140	ACGTGCTTCGCGACGTGGCGAGCGCCCTTGCCCTTGAGGCGCGCGCCCTTGA	199
QY	181	TCCCGCGGCGGTTGCGCGAGTGGGTGTCGTCCTTGAGCATGCGCGCATTCATTC	240
DB	200	TCCCGCGGCGGTTGCGCGAGTGGGTGTCGTCCTTGAGCATGCGCGCATTCATTC	259
QY	241	AGCCGCGAGAGACTAGCTCCGACCCCGCGCGCGCGGCTTCCAGAGGTTCACTG	300
DB	260	AGCCGCGAGAGACTAGCTCCGACCCCGCGCGCGCGGCTTCCAGAGGTTCACTG	319
QY	301	CTTCCTTCCCTGACTGACGCGCCCAATTACAGCAAAAGCTTGAAGCTTGACGCGCACCTGTG	360
DB	320	CTTCCTTCCCTGACTGACGCGCCCAATTACAGCAAAAGCTTGAAGCTTGAAGCGCACCTGTG	379
QY	361	CAAGCAGACGGGGGAGAGAACCAATTTGTGTGACTATGAAGGTTGTGGAAGGCTTCAAT	420
DB	380	CAAGCAGACGGGGGAGAGAACCAATTTGTGTGACTATGAAGGTTGTGGAAGGCTTCAAT	439
QY	421	CAGGACTACCATCTGAGACCGGCACATTTCTGACTACACAGAGAAAACCGCTTGTGG	480
DB	440	CAGGACTACCATCTGAGACCGGCACATTTCTGACTACACAGAGAAAACCGCTTGTGG	499
QY	481	TGCAGCCACTGGCTGTGATCAAAAAATTCACAACAATAATCAAACTTGAAGAAACATTTTGA	540
DB	500	TGCAGCCAAATGGCTGTGATCAAAAAATTCACAACAATAATCAAACTTGAAGAAACATTTTGA	559
QY	541	ACGCAAAATGAAAAATCAACAAAAAACAATATATATGCAAGCTTTTGAAGACTGTAAAGAC	600
DB	560	ACGCAAAATGAAAAATCAACAAAAAACAATATATATGCAAGCTTTTGAAGACTGTAAAGAC	619
QY	601	CTTTAAGAAACATCGACAGCTGAAAAATCATAGTGCACAGCTACCAATGAAGCTCTATT	660
DB	620	CTTTAAGAAACATCGACAGCTGAAAAATCATAGTGCACAGCTACCAATGAAGCTCTATT	679

```
QY 661 CAAGTGTACCCAGGAGATGTGGGAAACACTTGTGATCAACCCAGCAAGCTGAACGACA 720
    |||||||
Db 680 CAACTGTACCCAGGAGATGTGGGAAACACTTGTGATCAACCCAGCAAGCTGAACGACA 739
QY 721 TGGCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 780
    |||||||
Db 740 TGGCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 799
QY 781 GACGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG 840
    |||||||
Db 800 GACGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG 859
QY 841 CCGGAAACATTTAAAGCAAAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGA 900
    |||||||
Db 860 CCGGAAACATTTAAAGCAAAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGA 919
QY 901 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGAACCTATATCTGTGTTAA 960
    |||||||
Db 920 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGAACCTATATCTGTGTTAA 979
QY 961 TCTCAAAAGCATATCTCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020
    |||||||
Db 980 TCTCAAAAGCATATCTCTCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1039
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGTCTGACATAGCATGCTGTGTACATGA 1080
    |||||||
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAACAAAGTCTGACATAGCATGCTGTGTACATGA 1099
QY 1081 TCTGTACAAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTAAAAAGGAGTTTGGC 1140
    |||||||
Db 1100 TCTGTACAAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTAAAAAGGAGTTTGGC 1159
QY 1141 CTCTCATCTCGATGATATATCCCTCCCAAAAGGAAACAGGCAAGCTTATCTTTGTG 1200
    |||||||
Db 1160 CTCTCATCTCGATGATATATCCCTCCCAAAAGGAAACAGGCAAGCTTATCTTTGTG 1219
QY 1201 TCAAAAGGAGAGTCAACCACTGTGTGAAGACAGATGCTCTGACATTTGCAAGTACT 1260
    |||||||
Db 1220 TCAAAAGGAGAGTCAACCACTGTGTGAAGACAGATGCTCTGACATTTGCAAGTACT 1279
QY 1261 TACCTTTGGCTAA 1273
    |||||||
Db 1280 TACCTTTGGCTAA 1292

RESULT 11
US-09-724-676-44623
; Sequence 44623, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676
; FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44623
; LENGTH: 2352
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-44623

Query Match 99.6%; Score 1268.2; DB 5; Length 2352;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 121 ACGTGTCTGGCAGTGGCAGCGGCTTGCCCTGGGCTTGGAGGCGCGGCGCTTGGA 180
    |||||||
Db 140 ACGTGTCTGGCAGTGGCAGCGGCTTGCCCTGGGCTTGGAGGCGCGGCGCTTGGA 199
QY 181 TCGCGCGCGGTGGTGGCGAGTGGTGTGCTCTTGACCATGCGCGAGCGTTTCAATGC 240
    |||||||
Db 200 TCGCGCGCGGTGGTGGCGAGTGGTGTGCTCTTGACCATGCGCGAGCGTTTCAATGC 259
QY 241 AGCGGCGGAAGCTCAGCTCCGACCCCGCGCGCGCGCTTCCAGAGAGTTCAATCTG 300
    |||||||
Db 260 AGCGGCGGAAGCTCAGCTCCGACCCCGCGCGCGCGCTTCCAGAGAGTTCAATCTG 319
QY 301 CTCTTCCCTGACGTGCGAGCGCAATTTACAGCAAGCGCTGAGACCTTGACCGACCTGTG 360
    |||||||
Db 320 CTCTTCCCTGACGTGCGAGCGCAATTTACAGCAAGCGCTGAGACCTTGACCGACCTGTG 379
QY 361 CAAGCAGCGGGGAGAGACATTTGTTGTGACTATGAAGGCTGTGGCAAGGCTTCAT 420
    |||||||
Db 380 CAAGCAGCGGGGAGAGACATTTGTTGTGACTATGAAGGCTGTGGCAAGGCTTCAT 439
QY 421 CAGGAGTACCATCTGAGCGGCCACATTTGACTCAGCAGAGAGAGAGAGAGCGTTTGTG 480
    |||||||
Db 440 CAGGAGTACCATCTGAGCGGCCACATTTGACTCAGCAGAGAGAGAGAGAGCGTTTGTG 499
QY 481 TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATCAAACTGGAAGAAACATTTTGA 540
    |||||||
Db 500 TGCAGCCAAATGGCTGTATCAAAAATTCACACAAAATCAAACTGGAAGAAACATTTTGA 559
QY 541 ACGCAAAACATGAATAATCAACAAAACAAATATATATGCAATTTTGAAGACTGTAAAGAC 600
    |||||||
Db 560 ACGCAAAACATGAATAATCAACAAAACAAATATATATGCAATTTTGAAGACTGTAAAGAC 619
QY 601 CTTTAAAGAAATCAGCAGCTGAAGTCAATTCAGTGGCAGCATTCAGATTAACCTCTATT 660
    |||||||
Db 620 CTTTAAAGAAATCAGCAGCTGAAGTCAATTCAGTGGCAGCATTCAGATTAACCTCTATT 679
QY 661 CAACTGTACCCAGGAGATGTGGGAAACACTTGTGATCAACCCAGCAAGCTGAAGAGACA 720
    |||||||
Db 680 CAACTGTACCCAGGAGATGTGGGAAACACTTGTGATCAACCCAGCAAGCTGAAGAGACA 739
QY 721 TGGCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 780
    |||||||
Db 740 TGGCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 799
QY 781 GACGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG 840
    |||||||
Db 800 GACGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG 859
QY 841 CCGGAAACATTTAAAGCAAAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGA 900
    |||||||
Db 860 CCGGAAACATTTAAAGCAAAAGATTACCTTAAGCAACATGAAATCTATGCCCCAGA 919
QY 901 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGACCTATATCTGTGTTAA 960
    |||||||
Db 920 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGACCTATATCTGTGTTAA 979
QY 961 TCTGTCAAAAGCATATCTCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020
    |||||||
Db 980 TCTGTCAAAAGCATATCTCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1039
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGTCTGACATAGCATGCTGTGTACATGA 1080
    |||||||
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAACAAAGTCTGACATAGCATGCTGTGTACATGA 1099
QY 1081 TCTGTACAAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTAAAAAGGAGTTTGGC 1140
    |||||||
Db 1100 TCTGTACAAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTAAAAAGGAGTTTGGC 1159
QY 1141 CTCTCATCTCGATGATATATCCCTCCCAAAAGGAAACAGGCAAGCTTATCTTTGTG 1200
    |||||||
Db 1160 CTCTCATCTCGATGATATATCCCTCCCAAAAGGAAACAGGCAAGCTTATCTTTGTG 1219
```

QY	1201	TCAAAACGGAGAGTGATACCCAACTGTGTGGAAACAAAGATGCTCTCGACAGTTGCAGTACT	1260
Db	1220	TCAAAACGGAGAGTGATACCCAACTGTGTGGAAACAAAGATGCTCTCGACAGTTGCAGTACT	1279
QY	1261	TACCCTTGCGCTAA 1273	
Db	1280	TACCCTTGCGCTAA 1292	

RESULT 12

US-09-724-676A-44623

; Sequence 44623, Application us/09724676A

; GENERAL INFORMATION:

; APPLICANT: CompuGen LTD

; TITLE OF INVENTION: Variants of alternative splicing

; FILE REFERENCE: 129181.4 CompuGen

; CURRENT APPLICATION NUMBER: US/09/724,676A

; CURRENT FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 97222

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 44623

; LENGTH: 2352

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-724-676A-44623

Query Match	99.6%	Score 1268.2	DB 5	Length 2352
Best Local Similarly	99.8%	Pred. No. 0		
Matches 1270	0	Mismatches 3	Indels 0	Gaps 0

QY	1	ATGCGCACAGGCGGGCGAGCGGGGGCGGTCTCTGTGATACCGCGCGCTCCGGAAAT	60
Db	20	ATGCCACAGAGGGGCGCCGACCGCGGGGCGGTCTCTGTGTATCCGCGCGCTCCGGAAAT	79
QY	61	GTGCCGCGGTCCGCGCGAAAGTTTCAAGCAGGAGCCGTGGGCGGGCGCGCTTCCGGC	120
Db	80	GTCCCGCGCGCGCGCGGAAGTTTCAACAGGAGCCGTGGGCGGGCGCGCGCTTCCGGC	139
QY	121	ACGTCTCTCGGACGTGGCAGCGCGGCTGTGGGCTTGGAGGCGCGCGCGGCTTGA	180
Db	140	ACGTCTCTCTCGACGTGGCAGCGCGGCTTGGAGGCGCGCGCGGCTTGA	199
QY	181	TCCGCGCGCGCGGTGTGCGCGAGTCCGGTGTCTTGTGACATCGCGGAGCGCTTATTC	240
Db	200	TCCGCGCGCGCGGTGTGCGCGAGTCCGTGTGCTTGTGACATCGCGGAGCGCTTATTC	259
QY	241	AGCGGCGCAGAGCTAGTCTCGAGCCCGCGCGCGCGCTTCCAGAGGTTTATCTG	300
Db	260	AGCGGCGCAGAGCTAGTCTCGAGCCCGCGCGCGCGCTTCCAGAGGTTTATCTG	319
QY	301	CTCCTTCCCTGACTGACAGCGCCATTATACAGCAAGCCTGGAAAGTTTGAAGGCGACCTGG	360
Db	320	CTCCTTCCCTGACTGACAGCGCCATTATACAGCAAGCCTGGAAAGTTTGAAGGCGACCTGG	379
QY	361	CAAGCACAAGGGGGAGAGACCAATTTGTTGTGACTATGAAGGGTGTGGCAAGGCTTCAT	420
Db	380	CAAGCACAAGGGGGAGAGACCAATTTGTTGTGACTATGAAGGGTGTGGCAAGGCTTCAT	439
QY	421	CAGGACATNCCATCTGAGACCGCGCACATTTCTGACTACACAGGAGAAAGCCGTGTTTG	480
Db	440	CAGGACATNCCATCTGAGACCGCGCACATTTCTGACTACACAGGAGAAAGCCGTGTTTG	499
QY	481	TGCAGCCACTGGCTGTGATCAAAATTTCAACAGCAAAATTCAAACTTGAAGCAAACTTTTGA	540
Db	500	TGCAGCCAAATGGCTGTGATCAAAATTTCAACAGCAAAATTCAAACTTGAAGCAAACTTTTGA	559
QY	541	ACGCAACATGAAATCAACAAAAACAATATATATGACGTTTGAAGACTGTAGAAGAC	600
Db	560	ACGCAACATGAAATCAACAAAAACAATATATATGACGTTTGAAGACTGTAGAAGAC	619
QY	601	CTTTTAAGAAATCATGCAAGCTCAAAATTCATCAGTGCAGCATTACCATGAACCTCTAT	660
Db	620	CTTTTAAGAAATCATGCAAGCTCAAAATTCATCAGTGCAGCATTACCATGAACCTCTAT	679

QY	661	CAAGGTGTCCCGAGGAAGATGTGGGAAACACTTTTGATCATTCCACCGACGAAGCTGAAACGCA	720
Db	680	CAAGTGTACCCAGCAGAGATGTGGGAAACACTTTTGATCATTCCACCGACGAAGCTGAAACGCA	739
QY	721	TGCGCAAGGCCACGACGAGGCTATGTATGTCAAAAAGATGTTCCTTGTGTGGCAAAAACATG	780
Db	740	TGCGCAAGGCCACGACGAGGCTATGTATGTCAAAAAGATGTTCCTTGTGTGGCAAAAACATG	799
QY	781	GACGGAACTTCTGAAACACTGTGAGAGAAACCCCTTAAGAGAGAAATTACTATGTGAAGATG	840
Db	800	GACGGAACTTCTGAAACACTGTGAGAGAAACCCCTTAAGAGAGAAATTACTATGTGAAGATG	859
QY	841	CCGGAAACCATTTTAAACCCAAAGATTACCTTAAGCAACACATGAGAAACTCATGCCCCAGA	900
Db	860	CCGGAAACCATTTTAAACCCAAAGATTACCTTAAGCAACACATGAGAAACTCATGCCCCAGA	919
QY	901	AAGGATGTATGTGCGTGTGCCAAGAGAGCGCTGTGGAAGAACCTATACTACTGTGTTAA	960
Db	920	AAGGATGTATGTGCGTGTGCCAAGAGAGCGCTGTGGAAGAACCTATACTACTGTGTTAA	979
QY	961	TCTCCAAAGCCATATCCTCTCTCTTCATGAGAGAAAGCCGCCCTTTTGTGTGTAACATGC	1020
Db	980	TCTCCAAAGCCATATCCTCTCTCTTCATGAGAGAAAGCCGCCCTTTTGTGTGTAACATGC	1039
QY	1021	TGGCTGTGGCAAAACATTTTGGCAATGAAGAAACAAGTCTCACTAGCGATGCTGTTGTACATGA	1080
Db	1040	TGGCTGTGGCAAAACATTTTGGCAATGAAGAAACAAGTCTCACTAGCGATGCTGTTGTACATGA	1099
QY	1081	TCTGTACAAAGAAATAATGAAGCTCAAAAGTCAAAAATCTCTGTGAAAAACGAGCTTTGGC	1140
Db	1100	TCTGTACAAAGAAATAATGAAGCTCAAAAGTCAAAAATCTCTGTGAAAAACGAGCTTTGGC	1159
QY	1141	CTCTCATGTCTGATGGATTTATCCCTCCCAAAAAGAAACAAGGGCAAGGCTTATGTTTGTG	1200
Db	1160	CTCTCATGTCTGATGGATTTATCCCTCCCAAAAAGAAACAAGGGCAAGGCTTATGTTTGTG	1219
QY	1201	TCAAAACGAGAGTCAACCAACTGTGTGGAAGACAGAATGCTCTCGACAGTTGGACGACT	1260
Db	1220	TCAAAACGAGAGTCAACCAACTGTGTGGAAGACAGAATGCTCTCGACAGTTGGACGACT	1279
QY	1261	TACCGTTGGCTAA 1273	
Db	1280	TACCGTTGGCTAA 1292	

RESULT 13
US-09-724-676-44601
; Sequence 44601, Application US/09724676

```

; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing

```

; CURRENT APPLICATION NUMBER: US/09/724,676
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44601
; LENGTH: 2384
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-44601

```

Query Match	99.6%	Score 1268.2	DB 5	Length 2384
Best Local Similarity	99.8%	Pred. NO. 0		
Matches 1270; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY 1 ATGGCGACACACGGGCGCCGACGCGGGGGGCGTCTGATGACCCGGCGCGCTCCCGGAAAGT 60
Db 20 ATGGCGACACACGGGCGCCGACGCGGGGGCGGTGCTGTGTGACCGCGCGCGCTCCCGGAAAGT 79
QY 61 GTGGCGGCGTCCGCGGAAAGCTTCAACAGGAGCCGTGGGGCGGCGCGCCGCTTCCCGGC 120

Db 80 GTCCGGGCGCCGCGAAGTTTCAGCAGGAGCCGTGGCCGCGCGCTTCCGGC 139
QY 121 ACGTGTCTGGCAGCTGGCAGCGCGCTGGGCTTGGAGGCGCGCCGCTTGA 180
Db 140 ACGGTCTTGGCAGCTGGCAGCGCGCTGGGCTTGGAGGCGCGCCGCTTGA 199
QY 181 TCCGCGCGCGTGGTCCCGCAGTGGTGGTCTCTTGAACATCCGACGCGTTTCA 240
Db 200 TCCGCGCGCGTGGTCCCGCAGTGGTGGTCTCTTGAACATCCGACGCGTTTCA 259
QY 241 AGCCGGGAGAGCTCACTCCGACCCCGCGCGCGCTTCCAGAGGTTCACTG 300
Db 260 AGCCGGGAGAGCTCACTCCGACCCCGCGCGCGCTTCCAGAGGTTCACTG 319
QY 301 CTCCTTCCCTGACTGACGCGCAATTAAGCAAGCCTGGAAGTTGACGCGCACTGTG 360
Db 320 CTCCTTCCCTGACTGACGCGCAATTAAGCAAGCCTGGAAGTTGACGCGCACTGTG 379
QY 361 CAACGACACGGGGGAGAGACATTTGTTGTACTATGAAGGTTGGCAAGCCCTTCA 420
Db 380 CAACGACACGGGGGAGAGACATTTGTTGTACTATGAAGGTTGGCAAGCCCTTCA 439
QY 421 CAGGACTACCATCTGAGCGCCCATTTGACTCAGACAGGAGAAAGCCGTTTGTG 480
Db 440 CAGGACTACCATCTGAGCGCCCATTTGACTCAGACAGGAGAAAGCCGTTTGTG 499
QY 481 TGCAGCCTGCTGTATCAAAATTCACACAAATTCAACTGTAAGAACTTTTGA 540
Db 500 TGCAGCCTGCTGTATCAAAATTCACACAAATTCAACTGTAAGAACTTTTGA 559
QY 541 ACGCAACATGAAATATCAACAAATATATATGAGTTTGAAGCTGTAGAAGAC 600
Db 560 ACGCAACATGAAATATCAACAAATATATATGAGTTTGAAGCTGTAGAAGAC 619
QY 601 CTTTAAAGACATGACAGCTGAAATATCATGATGCCAGATACCAATGAACTTAT 660
Db 620 CTTTAAAGACATGACAGCTGAAATATCATGATGCCAGATACCAATGAACTTAT 679
QY 661 CAATGTACCCAGGAGAGATGTGGAAACATTTGATCACCCAGCAAGCTGAAGCA 720
Db 680 CAATGTACCCAGGAGAGATGTGGAAACATTTGATCACCCAGCAAGCTGAAGCA 739
QY 721 TGCAGGCGCCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGGCAAAAGATG 780
Db 740 TGCAGGCGCCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGGCAAAAGATG 799
QY 781 GACGGAATCTGTGAACATGTGAGAGAAAGCCATTAAGAGAAATACTATGTAAGTATG 840
Db 800 GACGGAATCTGTGAACATGTGAGAGAAAGCCATTAAGAGAAATACTATGTAAGTATG 859
QY 841 CCGGAAACATTTTAAAGCAGAAATTAACCTTAACACACATGAAAGCTCATGGCCGGA 900
Db 860 CCGGAAACATTTTAAAGCAGAAATTAACCTTAACACACATGAAAGCTCATGGCCGGA 919
QY 901 AAGGATGTATGTGCTGTCCAGAGAGAGCTGTGAAGAACTTAATCTGTTTAA 960
Db 920 AAGGATGTATGTGCTGTCCAGAGAGAGCTGTGAAGAACTTAATCTGTTTAA 979
QY 961 TCTCCAAAGCATATCTCTCTCTTCCATGAGAGAAAGCCCTTTTGTGTGAACATGC 1020
Db 980 TCTCCAAAGCATATCTCTCTCTTCCATGAGAGAAAGCCCTTTTGTGTGAACATGC 1039
QY 1021 TGGGTGTGGCAAAATTTGATGAAGAAAGTCTCACTAGGATGCTGTGAACATGA 1080
Db 1040 TGGGTGTGGCAAAATTTGATGAAGAAAGTCTCACTAGGATGCTGTGAACATGA 1099
QY 1081 TCTGTGACAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTGAAGAAAGCAGTGTGGC 1140
Db 1100 TCTGTGACAGAGAAATGAAGCTCAAGTCAAAAATCTCTGTGAAGAAAGCAGTGTGGC 1159
QY 1141 CTCCTCATCTGATGATATATCCCTCCCAAGAGAAAGCAAGGCTTATCTTTGTG 1200
Db 1160 CTCCTCATCTGATGATATATCCCTCCCAAGAGAAAGCAAGGCTTATCTTTGTG 1219

QY 1201 TCAGAAAGAGAGTACACCAACTGTGTGAAGAGCAAGATGCTGTGACAGTTGCACT 1260
Db 1220 TCAGAAAGAGAGTACACCAACTGTGTGAAGAGCAAGATGCTGTGACAGTTGCACT 1279
QY 1261 TACCTTGGCTAA 1273
Db 1280 TACCTTGGCTAA 1292

RESULT 14
US-09-724-676A-44601
; Sequence 44601, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44601
; LENGTH: 2384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-44601

Query Match 99.8%; Score 1268.2; DB 5; Length 2384;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGCGACAGCGCGCGCGAGCGCGCGCGTCCGTGTGACCGCGCGCGTCCCGAGT 60
Db 20 ATGGCGACAGCGCGCGCGCGAGCGCGCGCGTCCGTGTGACCGCGCGCGTCCCGAGT 79
QY 61 GTGCGGCGCTGCGCGCGAAGTTTCAAGAGAGCGCGCGCGCGCGCGCGCGTCCCGG 120
Db 80 GTGCGGCGCGCGCGCGCGAAGTTTCAAGAGAGCGCGCGCGCGCGCGCGCGTCCCGG 139
QY 121 ACGTGTCTGGCAGCTGAGCGCGCGCGCTGCGCGCTTGGAGGCGCGCGCGCGCTGGA 180
Db 140 ACGTGTCTGGCAGCTGAGCGCGCGCGCTGCGCGCTTGGAGGCGCGCGCGCGCTGGA 199
QY 181 TCCGCGCGCGTGGTCCCGCAGTGGTGGTCTCTTGAACATCCGACGCGTTTCA 240
Db 200 TCCGCGCGCGTGGTCCCGCAGTGGTGGTCTCTTGAACATCCGACGCGTTTCA 259
QY 241 AGCCGGGAGAGCTCACTCCGACCCCGCGCGCGCGCTTCCAGAGGTTCACTG 300
Db 260 AGCCGGGAGAGCTCACTCCGACCCCGCGCGCGCGCTTCCAGAGGTTCACTG 319
QY 301 CTCCTTCCCTGACTGACGCGCAATTAAGCAAGCCTGGAAGCTTACGCGCACTGTG 360
Db 320 CTCCTTCCCTGACTGACGCGCAATTAAGCAAGCCTGGAAGCTTACGCGCACTGTG 379
QY 361 CAACGACACGGGGGAGAGACATTTTGTGACTATGAAGGTTGGCAAGGCTTCA 420
Db 380 CAACGACACGGGGGAGAGACATTTTGTGACTATGAAGGTTGGCAAGGCTTCA 439
QY 421 CAGGACTACCATCTGAGCGCCCATTTGACTCAGACAGGAGAAAGCCGTTTGTG 480
Db 440 CAGGACTACCATCTGAGCGCCCATTTGACTCAGACAGGAGAAAGCCGTTTGTG 499
QY 481 TGCAGCCTGCTGTATCAAAATTCACACAAATTAATCTGAAGAAAGCAGTGTGGC 540
Db 500 TGCAGCCTGCTGTATCAAAATTCACACAAATTAATCTGAAGAAAGCAGTGTGGC 559
QY 541 ACGCAACATGAAATATCAACAAATATATATGAGTTTGAAGAGTGAAGAGAC 600
Db 560 ACGCAACATGAAATATCAACAAATATATATGAGTTTGAAGAGTGAAGAGAC 619
QY 601 CTTTAAAGACATGACAGCTGAAATCATCACTGACAGATACCAATGAACTTAT 660


```

Db 1160 CTCATCTCAGTGGATATATCCCTCCCAAGGAAACAAGGCAAGCTTATCTTGTG 1219
QY 1201 TCAAAACGAGAGTCAACCAACTGTGTGSAAGACAAGATGCTCTCGACAGTGCAGTACT 1260
    |||||
Db 1220 TCAAAACGAGAGTCAACCAACTGTGTGSAAGACAAGATGCTCTCGACAGTGCAGTACT 1279
QY 1261 TACCTTGGCTAA 1273
    |||||
Db 1280 TACCTTGGCTAA 1292

```

Search completed: February 10, 2003, 13:06:19
 Job time : 242.015 secs

THIS PAGE BLANK (USPTO)

TITLE Fujiiwara, T., Takahashi, E., Shin, S. and Nakamura, Y.
Molecular cloning, characterization, and chromosomal mapping of a novel human gene (GTF3A) that is highly homologous to Xenopus transcription factor Iiia
JOURNAL Cytogenet. Cell Genet. 70 (3-4), 235-238 (1995)
MEDLINE 95309028
REFERENCE 2 (bases 1 to 1381)
AUTHORS Nakamura, Y.
TITLE Direct Submision
JOURNAL Submitted (22-JUL-1994) Yusuke Nakamura, Cancer Institute, Department of Biochemistry, 1-37-1 Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan (E-mail: nakamura@ganvxl.jicr.or.jp, Tel: 03-3918-0111 (ex.4501), Fax: 03-3918-0342)
location/Qualifiers
1. 1381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="Library of T. Fujiiwara, S. Shin and Y. Nakamura"
1. 1381
/gene="GTF3A"
20. 1291
/gene="GTF3A"
/codon_start=1
/product="Xenopus transcription factor Iiia homologue"
/protein_id="BA06988.1"
/db_xref="GI:1616942"
/translation="MRSQADAGRCILVTARQSVNPSRSGSRGAPRPARVSA KQSAKPGIAGCAGADPRPAVVAESVSLITADAFIAGSSSAPTPRPRLPRFICSPDSANYSKAMKLDALHCKHTGERPVCYEGGKAFIDYHLIRLHTREKPRVC AANGCQKFNKTSNLKHEKHEHNOQYICSEDEKTFPKHQQKLIHQCONTREP LKNCIOEGCGKHFASPSKLRHAKAHEGYVQCGCFVAKWTLELKHVETHEIL CEVCRTEPKRKIDILKOHMKTTHAERDVQCPREGCRRYTYVNLQSHLSFHEESRP FYCEHAGCGKTFAMKQSLTRHAVVHDPPDKKKLKKRREKREBELSSOMTYPKPK KQCGSLICQNGESPNCVEDKMLSTVAVLTLG"
BASE COUNT 386 a 350 c 354 g 291 t
ORIGIN
Query Match 96.9%; Score 1233.2; DB 9; Length 1381;
Best Local Similarity 99.5%; Pred. No. 4.1e-259;
Matches 1268; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 1 ATGGCGACAGACGGCGCGCGGGGGGGTGGCTGGTGACCGCGCGCGCTCCCGGAAGT 60
|||||
Db 20 ATGCGCGACAGCGCGCGCGCGGGGGGGTGGCTGGTGACCGCGCGCGCTCCCGGAAGT 79
QY 61 GTGCGCGCGCGCGCGCGGAAGTTTCAGCAGGAGAGCGCTGGGGCGCGCGCGCTCCCGGC 120
|||||
Db 80 GTGCGCGCGCGCGCGCGGAAGTTTCAGCAGGAGAGCGCTGGGGCGCGCGCGCTCCCGGC 138
QY 121 AGCTGTCTGGGACAGTGGCAGCGCGCTGGCCCTGGGCTTGAAGGCGCGCGCGCTTGA 180
|||||
Db 139 AGCTGTCTGGGACAGTGGCAGCGCGCTGGCCCTGGGCTTGAAGGCGCGCGCGCTTGA 198
QY 181 TCGCGCGCGCGCGCGCGAGTGGGTGGTGGTGGTGGACCATGCGCGACCGCTTCATGGC 240
|||||
Db 199 TCGCGCGCGCGCGCGCGAGTGGGTGGTGGTGGTGGACCATGCGCGACCGCTTCATGGC 258
QY 241 AGCGCGCGAGAGTCAAGTCTCGACCCCGCGCGCGCGCGCTTCCAGAGAGTTTCACTG 300
|||||
Db 259 AGCGCGCGAGAGTCAAGTCTCGACCCCGCGCGCGCGCGCTTCCAGAGAGTTTCACTG 318
QY 301 CTGCTTCCCTGAGTGCAGCGCGCAATTACAGAAAGCCTGGAAGCTTGAAGCGGACCTGTG 360
|||||
Db 319 CTGCTTCCCTGAGTGCAGCGCGCAATTACAGAAAGCCTGGAAGCTTGAAGCGGACCTGTG 378
QY 361 CAAGCACACGCGGGAGAGACATTTGTTGTGACTATGAAGGTTGTGCGAAGCGCTTCAT 420
|||||
Db 379 CAAGCACACGCGGGAGAGACATTTGTTGTGACTATGAAGGTTGTGCGAAGCGCTTCAT 438
QY 421 CAGGAGTACCATCTGAGCGCGCAATTTGACTCACACAGGAGAAAGCGTTTGTGTTG 480
|||||
Db 439 CAGGAGTACCATCTGAGCGCGCAATTTGACTCACACAGGAGAAAGCGTTTGTGTTG 498

QY 481 TGCAGCCACTGGCTGTGATCAAAAAATTCAACACAAAAATCAAACTGGAAGAAACATTTTGA 540
|||||
Db 499 TGCAGCCAAATGGCTGTGATCAAAAAATTCAACACAAAAATCAAACTGGAAGAAACATTTTGA 558
QY 541 ACAGCAACATGAAAAATCAACAAAAACAATATATATGACGTTTGAAGACGTGAAGAGAC 600
|||||
Db 559 ACCGAAACATGAAAAATCAACAAAAACAATATATATGACGTTTGAAGACGTGAAGAGAC 618
QY 601 CTTTAGAAAAATCAGCAGCTGAAATTCATCAGTGTCCAGCATACCAATGAACCTCTATT 660
|||||
Db 619 CTTTAAACAAATCAGCAGCTGAAATTCATCAGTGTCCAGCAATACCAATGAACCTCTATT 678
QY 661 CAAGTGTACCCAGAAAGATGTGGGAAACACTTTCATCCACCGCAAGGTGAAGAGACA 720
|||||
Db 679 CAAGTGTACCCAGAAAGATGTGGGAAACACTTTCATCCACCGCAAGGTGAAGAGACA 738
QY 721 TGCCAAAGCCACGAGGCGTATGTATGTCAAAAAAGATGTTCTTGTGGCAAAAAACATG 780
|||||
Db 739 TGCCAAAGCCACGAGGCGTATGTATGTCAAAAAAGATGTTCTTGTGGCAAAAAACATG 798
QY 781 GAGCGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACATATGTGAACTATG 840
|||||
Db 799 GAGCGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACATATGTGAACTATG 858
QY 841 CCGGAAACATTTAAACGCAAAAGATTACCTTAAAGCAACACATGAAAACTCATGCCCCAGA 900
|||||
Db 859 CCGGAAACATTTAAACGCAAAAGATTACCTTAAAGCAACACATGAAAACTCATGCCCCAGA 918
QY 901 AAGGATGATGTGCGCTGTCCAAGAGAGCGTGTGAAGAAGCTATATCTGTGTTTAA 960
|||||
Db 919 AAGGATGATGTGCGCTGTCCAAGAGAGCGTGTGAAGAAGCTATATCAAGCTGTGTTTAA 978
QY 961 TCTCCAAAGCATATCTCTCTCTTCCATGAGGAAAGCGCCCTTTTGTGTGAACATGC 1020
|||||
Db 979 TCTCCAAAGCATATCTCTCTCTTCCATGAGGAAAGCGCCCTTTTGTGTGAACATGC 1038
QY 1021 TGCGCTGTGCAAAACATTTGCAATGAAACAAAGTCATCAGGCAATGCTTGTATACATGA 1080
|||||
Db 1039 TGCGCTGTGCAAAACATTTGCAATGAAACAAAGTCATCAGGCAATGCTTGTATACATGA 1098
QY 1081 TCCTGACAGAGAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAATAAC-GGAGTTTGG 1139
|||||
Db 1099 TCCTGACAGAGAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAATAAC-GGAGTTTGG 1158
QY 1140 CCTCTCATCTCATGTGATATATCCCTCCCAAAAGAAACAAGGCAAGCTTATCTTTGT 1199
|||||
Db 1159 CCTCTCATCTCATGTGATATAT-CTTCCCAAAAGAAACAAGGCAAGCTTATCTTTGT 1217
QY 1200 GTCAAAACGAGAGTCAACCCCACTGTGTGAAGACAAAGATGCTCTGACAGTTGACAGTAC 1259
|||||
Db 1218 GTCAAAACGAGAGTCAACCCCACTGTGTGAAGACAAAGATGCTCTGACAGTTGACAGTAC 1277
QY 1260 TTACCCCTTGGCTAA 1273
|||||
Db 1278 TTACCCCTTGGCTAA 1291
RESULT 2
E10962
LOCUS E10962 1399 bp RNA linear PAT 29-SEP-1997
DEFINITION CDNA's encoding hTfIIa protein.
ACCESSION E10962
VERSION E10962.1 GI:22028822
KEYWORDS JP 1996070870-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1399)
AUTHORS Fujiiwara, T., Takeda, S., Shimada, Y., Ozaki, K. and Shin, T.
TITLE HUMAN TRANSCRIPTIONAL FACTOR Iiia GENE
JOURNAL Patent: JP 1996070870-A 1 19-MAR-1996;

COMMENT	OTSUKA PHARMACEUT CO LTD OS Homo sapiens (human) PN JP 19960708/0-4/1 PD 19-MAR-1996 PF 05-SEP-1994 JP 1994211022 PI FUJIMARA TSUTOMU, TAKEDA SEI, SHIMADA YOSHIKAZU, OZAKI KOICHI, PI SHIN TEIKIN PC C12N15/09//A61K31/70,A61K48/00,C07H21/04; CC Strandedness: Single; CC topology: linear; FH Key FH Location/Qualifiers FT source FT 1..1399 FT /organism='Homo sapiens' FT 5'UTR FT 1..19 FT CDS FT 20..1288 FT /product='hTFPIIIA protein' FT 1289..1399. FT Location/Qualifiers FT 3'UTR FT 1..1399 FT /organism='Homo sapiens' FT /db_xref='taxon:9606' <th>BASE COUNT</th> <th>405 a 349 c 354 g 291 t</th> <th>ORIGIN</th>	BASE COUNT	405 a 349 c 354 g 291 t	ORIGIN
Query Match	96.7% Score 1231.6; DB 6; Length 1399;			
Best Local Similarity	99.5%; Pred. No. 9.2e-259;			
Matches 1267; Conservative	0; Mismatches 4; Indels 3; Gaps 3.			
1 ATGCGCAGACGAGCGGCGCCGACGCGGGGCGCGTGCCTGATGACCGCGCGCGCTCCGGAAAGT	60			
20 ATGCGCAGACGAGCGGCGCCGACGCGGGGCGGTGCTGATGACCGCGCGCGCTCCGGAAAGT	79			
61 GTGCGCGCGCTCCGCGGAGAGTTACAGAGGAGACCCGTGGCGCGCGCGCGCTTCCGGC	120			
80 GTGCGCGCGCTCCGCGGAGAGTTACAGAGGAGACCCGTGGCGCGCGCGCGCTTCCGGC	138			
121 AGGTCTTCGACAGTGGAGCGCGCGCTGGCGCTTGGAGCGCGCGCGCGCTTGA	180			
139 AGGTCTTCGACAGTGGAGCGCGCGCTGGCGCTTGGAGCGCGCGCGCGCTTGA	198			
181 TCCGCGCGCGGTGTGTGCGCGAGTGGGTGTGCTCTTGACCATGCGCGAGCGCTTATTTGC	240			
199 TCCGCGCGCGGTGTGTGCGCGAGTGGGTGTGCTCTTGACCATGCGCGAGCGCTTATTTGC	258			
241 ACCCGCGGAGACTCAGCTCCGAGACCCCGCGCGCGCGCGCTTCCAGAGGTTATCTG	300			
259 ACCCGCGGAGACTCAGCTCCGAGACCCCGCGCGCGCGCGCTTCCAGAGGTTATCTG	318			
301 CTCCTTCCCTGACTCAGAGGCGCAATTAACAGCAAGAGCTTGAAGCTTGAAGCGCACCTGTG	360			
319 CTCCTTCCCTGACTCAGAGGCGCAATTAACAGCAAGAGCTTGAAGCTTGAAGCGCACCTGTG	378			
361 CAAGCACACGCGGGGAGAGACCAATTTGTTGTGACTATGAAAGGCTGTGCGAAGCGCTTCA	420			
379 CAAGCACACGCGGGGAGAGACCAATTTGTTGTGACTATGAAAGGCTGTGCGAAGCGCTTCA	438			
421 CAGGAGCATACATTCGAGCGCGGCACATTTGACTCACAACAGAGAGAAACCCCTTTGTTTG	480			
439 CAGGAGCATACATTCGAGCGCGGCACATTTGACTCACAACAGAGAGAAACCCCTTTGTTTG	498			
481 TGCAGCACTGCTGTGATCAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGA	540			
499 TGCAGCACTGCTGTGATCAAAATTTCAACACAAATCAAACTTGAAGAAACATTTTGA	558			
541 ACGCAACATGAAAAATCAACAAACAAATATATATGCAAGCTTTGAAGACTGTAAGAGAC	600			
559 ACGCAACATGAAAAATCAACAAACAAATATATATGCAAGCTTTGAAGACTGTAAGAGAC	618			
601 CTTTAAAGAAACATCAGCAGCTGAAAAATCATAGGCGCGCATACCAATGAACCTCTATT	660			
619 CTTTAAAGAAACATCAGCAGCTGAAAAATCATAGGCGCGCATACCAATGAACCTCTATT	678			

QY	661	CAAGTGTACCCAGAGAGATGTGGAAACACTTTTGCATACCCAGACAAAGCTGAAACGCA	720
Db	679	CAGAGTACCCAGAGAGATGTGGAAACACTTTTGCATACCCAGACAAAGCTGAAACGCA	738
QY	721	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCCCTTGTGGCAAAAACATG	780
Db	739	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCCCTTGTGGCAAAAACATG	798
QY	781	GACGGAACCTTCTGAACAATGTGAGAAACCCTAATAAGAGAAATACTATGTGAAGTATG	840
Db	799	GACGGAACCTTCTGAACAATGTGAGAAACCCTAATAAGAGAAATACTATGTGAAGTATG	858
QY	841	CCGGAACCAATTTAAACGCAAGATTACCTTATAGCACATGAAACTCATGTCCCAGA	900
Db	859	CCGGAACCAATTTAAACGCAAGATTACCTTATAGCACATGAAACTCATGTCCCAGA	918
QY	901	AAGGATGATGTGCTGCTCTCCCAAGAAAGCGTGTGGAAGAACCTATACTATGTGTTAA	960
Db	919	AAGGATGATGTGCTGCTCTCCCAAGAAAGCGTGTGGAAGAACCTATACTATGTGTTAA	978
QY	961	TCTCCAAAGCCATATCTCTCTCCCTCCATGAGAAAGCCGCCCTTTGTGTGGAACATGC	1020
Db	979	TCTCCAAAGCCATATCTCTCTCCCTCCATGAGAAAGCCGCCCTTTGTGTGGAACATGC	1038
QY	1021	TGGCTGTGGCAAAACATTTTGCATGATAAACAAGTCTACATAGCATGCTGTGTACATGA	1080
Db	1039	TGGCTGTGGCAAAACATTTTGCATGATAAACAAGTCTACATAGCATGCTGTGTACATGA	1098
QY	1081	TGCTGACAAAGAAGAAATGAAGTCAAGTCAAAAATCTGTGTGAANAAC-GGACTTTGG	1139
Db	1099	TGCTGACAAAGAAGAAATGAAGTCAAGTCAAAAATCTGTGTGAANAACGGGAATTTGG	1158
QY	1140	CCTTCATCTCAGTGGATATATCCCTCCCAAAAGAAACAAGGCCAAGCTTATCTTGT	1199
Db	1159	CCTTCATCTCAGTGGATATAT-CTCCCAAAAGAAACAAGGCCAAGCTTATCTTGT	1217
QY	1200	GTCAAACGAGAGATCACCAACTGTGTGGAAGACACAGATGCTCTCGACAGTTGCAGTAC	1259
Db	1218	GTCAAACGAGAGATCACCAACTGTGTGGAAGACACAGATGCTCTCGACAGTTGCAGTAC	1277
QY	1260	TTACCCCTTGGCTAA 1273	
Db	1278	TTACCCCTTGGCTAA 1291	
RESULT 3			
LOCUS	HSU20272	1173 bp	mRNA linear PRI 26-SEP-1995
DEFINITION	Human DNA/RNA-binding protein mRNA, partial cds.		
ACCESSION	U20272		
VERSION	U20272.1	GI:644870	
KEYWORDS	.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	1 (sites)		
AUTHORS	Drew, P.D., Nigle, J.W., Canning, R.D., Ozato, K., Biddison, W.E. and Becker, K.G.		
TITLE	Cloning and expression analysis of a human cDNA homologous to Xenopus pRiIIA		
JOURNAL	Gene 159 (2), 215-218 (1995)		
MEDLINE	95347600		
PUBMED	7622052		
REFERENCE	2 (bases 1 to 1173)		
AUTHORS	Becker, K.G.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JAN-1995) Kevin G. Becker, Nat. Inst. of Neur. Diseases and Stroke/NIH, Neuroimmunology Branch, 9000 Rockville Pike, Bethesda, MD 20892, USA		
FEATURES	Location/Qualifiers		
source	1..1173		
	/organism="Homo sapiens"		

```
/db_xref="taxon:9606"
/clone="C2H2-34.10"
/sex="female"
/tissue_type="brain: hippocampus"
/clone_lib="Stratagene #936205"
/dev_stage="2 years old"
<!. .1092
/note="similar to Xenopus TFIIA encoded by Swissprot
Accession Number P03001"
/codon_start=1
/product="DNA/RNA-binding protein"
/db_xref="GI:644871"
/translacion="PPAVVAESVSLTIADAFIAGESAPTPRPALPRRIFCSFPD
CSANYSKAMKIDAHLCIKHTEGEPVCDYEGCGKAFIRDIHLRSOLIKTHGEKPEVCA
NCCDQKPTKNSLKKHFKERHNOOKOYICSFEDCKTKFKKHQOLIKHOCQHTNEPLF
KTOGCGCKHFPSPSKLRKHAHAGSYOCQKCSYVATKTYTLHLHYETHREELICE
VCRKTFKRNDYLRKHMTKHAPEKRDYCRCPREGGKTYTIVEMLOSHLSIFBESRPV
CEHAGCGTKFAMKQSLTRAVVHPDKKMKLKVKSREKRSLSHLSGYIPKRRKG
QGLSLCONCESFNCVEDKMLSTVAVLTLG"
misc_feature
118..184
/note="encodes zinc-finger"
misc_feature
208..274
/note="encodes zinc-finger"
misc_feature
298..367
/note="encodes zinc-finger"
misc_feature
394..460
/note="encodes zinc-finger"
misc_feature
484..550
/note="encodes zinc-finger"
misc_feature
565..631
/note="encodes zinc-finger"
misc_feature
649..709
/note="encodes zinc-finger"
misc_feature
736..805
/note="encodes zinc-finger"
misc_feature
829..895
/note="encodes zinc-finger"
misc_feature
829..895
/note="encodes zinc-finger"
BASE COUNT 363 a 283 c 265 g 262 t
ORIGIN
Query Match 85.5%; Score 1088.8; DB 9; Length 1173;
Best Local Similarity 99.8%; Pred. No. 1.5e-227;
Matches 1090; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 182 CCGCGGCGCGTGGTGGCGGAGTGGGTGCTCTTGACATCGCCGCGCTTCGAGAGCTTCATCTGC 241
DB 1 CCGCGGCGCGTGGTGGCGGAGTGGGTGCTCTTGACATCGCCGCGCTTCGAGAGCTTCATCTGC 60
QY 242 GCGGCGGAGAGTCACTCCGACCCGCGCGCGCGCGCTTCGAGAGCTTCATCTGC 301
DB 61 GCGGCGGAGAGTCACTCCGACCCGCGCGCGCGCGCGCTTCGAGAGCTTCATCTGC 120
QY 302 TCGTTCCTGACTGACAGCGCAATACAGCAAAAGCTGGAAGCTTGACGGCACCTGTGC 361
DB 121 TCGTTCCTGACTGACAGCGCAATACAGCAAAAGCTGGAAGCTTGACGGCACCTGTGC 180
QY 362 AAGCACGCGGAGAGACCATTTGTTTGACTATGAAGAGTGTGGCAAGCGCTTCATC 421
DB 181 AAGCACGCGGAGAGACCATTTGTTTGACTATGAAGAGTGTGGCAAGCGCTTCATC 240
QY 422 AGGAGTACCATCTGAGCCGCAATCTGACTACAGACAGAGAAAAAGCGTTTGTGT 481
DB 241 AGGAGTACCATCTGAGCCGCAATCTGACTACAGACAGAGAAAAAGCGTTTGTGT 300
QY 482 GCGAGCATGCTGCTGTATCAAAATTCACACAAATCAAACTTGAAGAACTTTGAA 541
DB 301 GCGAGCATGCTGCTGTATCAAAATTCACACAAATCAAACTTGAAGAACTTTGAA 360
QY 542 CGCAAAACATGAATAAACAATAATATATGAGTTTGAAGAGCTGTAGAAGACC 601
DB 361 CGCAAAACATGAATAAACAATAATATATGAGTTTGAAGAGCTGTAGAAGACC 420
```

```
QY 602 TTTAAGAACATCAGCAGCTGAAATTCATCAGTGGCAGCATACCAATGAACTCTATTG 661
DB 421 TTTAAGAACATCAGCAGCTGAAATTCATCAGTGGCAGCATACCAATGAACTCTATTG 480
QY 662 AAGTGTACCCAGAGAGATGTGGGAAACCTTTGCATCCCAAGCAAGCTGAAAGCAGT 721
DB 481 AAGTGTACCCAGAGAGATGTGGGAAACCTTTGCATCCCAAGCAAGCTGAAAGCAGT 540
QY 722 GCCAAGGCCAGAGAGGCTATGTATGTCAAAAAGATGTTCCTTTGTGGCAAAACATGG 781
DB 541 GCCAAGGCCAGAGAGGCTATGTATGTCAAAAAGATGTTCCTTTGTGGCAAAACATGG 600
QY 782 ACGGAACCTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATGC 841
DB 601 ACGGAACCTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATGC 660
QY 842 CGGAAACATTTTAACGCAAAATTAACCTTAAGCAACATGAAAGCTATATCCCGAGAA 901
DB 661 CGGAAACATTTTAACGCAAAATTAACCTTAAGCAACATGAAAGCTATATCCCGAGAA 720
QY 902 AGGATGTATGTGCTGTCTCAAGAGAGGCTGTGCAAGCACTATACCTATGTGTTTAA 961
DB 721 AGGATGTATGTGCTGTCTCAAGAGAGGCTGTGCAAGCACTATACCTATGTGTTTAA 780
QY 962 CTCCAAAAGCCATATCTCTCTCCATGAGAAAGCGCCCTTTGTGTGTAACATGCT 1021
DB 781 CTCCAAAAGCCATATCTCTCTCCATGAGAAAGCGCCCTTTGTGTGTAACATGCT 840
QY 1022 GGCTGTGGCAAAACATTTTGCATGAACAAGCTCTACAGCATGCTTGTATGATGAT 1081
DB 841 GGCTGTGGCAAAACATTTTGCATGAACAAGCTCTACAGCATGCTTGTATGATGAT 900
QY 1082 CCTGACAGAGAAATGAAGCTCAAAAGCTCAAAATTCGTGAAAGAGGATTTGGCC 1141
DB 901 CCTGACAGAGAAATGAAGCTCAAAAGCTCAAAATTCGTGAAAGAGGATTTGGCC 960
QY 1142 TCTCATCTCAGTGTATATTCCTCCCAAGGAAAGGAGGCAAGCTTATCTTTGTGT 1201
DB 961 TCTCATCTCAGTGTATATTCCTCCCAAGGAAAGGAGGCAAGCTTATCTTTGTGT 1020
QY 1202 CAAAGCGAGAGTCAACCACTGTGTGAAAGACAGATGCTCTCGACATTTGCAGTACT 1261
DB 1021 CAAAGCGAGAGTCAACCACTGTGTGAAAGACAGATGCTCTCGACATTTGCAGTACT 1080
QY 1262 ACCCTGGCTAA 1273
DB 1081 ACCCTGGCTAA 1092
RESULT 4
HSU14134 1098 bp mRNA linear PRI 01-OCT-1994
LOCUS Human transcription factor TFIIA (TFIIIA) mRNA, partial cds.
DEFINITION U14134
ACCESSION U14134
VERSION U14134.1 GI:551534
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1098)
Drew, P.D., Nagle, J.W., Canning, R.D., Ozato, K., Biddison, W.E. and
Becker, K.G.
Cloning and expression analysis of a cDNA encoding human TFIIA
TITLE Unpublished
REFERENCE 2 (bases 1 to 1098)
Becker, K.G.
Direct Submision
JOURNAL Submitted (29-AUG-1994) Kevin G. Becker, National Institute of
Neurological Diseases and Stroke/NIH, Neuroimmunology Branch, 9000
Rockville Pike, Bethesda, MD 20892, USA
FEATURES
source 1..1098
```

Query Match	Best Local Similarity	Score	DB	Length
Matches 100%; Conservative	92.4%; Pred. No. 1.7e-190;	0; Mismatches 8;	Indels 75;	Gaps 1;
182	CCGCGCGAGAGCTCAGTCCGACCCCGCGCGCCCGCGCTTCCGAGAGCTTCATTCGCA	241		
Db	1 CCGCGCGAGAGCTCAGTCCGACCCCGCGCGCCCGCGCTTCCGAGAGCTTCATTCGCA	60		
QY	242 GCGCGCGAGAGCTCAGTCCGACCCCGCGCGCCCGCGCTTCCGAGAGCTTCATTCGCA	301		
Db	61 GCGCGCGAGAGCTCAGTCCGACCCCGCGCGCCCGCGCTTCCGAGAGCTTCATTCGCA	120		
QY	302 TCCTTCCTGAGTGCAGCGCCCAATTACAGCAAAAGCGTGGAGCTTGAGCGGACCTGTGC	361		
Db	121 TCCTTCCTGAGTGCAGCGCCCAATTACAGCAAAAGCGTGGAGCTTGAGCGGACCTGTGC	180		
QY	362 AAGCACAGGCGGAGAGACATTTGTTTGTGACTATGAAAGGCTGTGGCAAGGCGTTCAATC	421		
Db	181 AAGCACAGGCGGAGAGACATTTGTTTGTGACTATGAAAGGCTGTGGCAAGGCGTTCAATC	240		
QY	422 AGGCACTACCACTGAGCGGCACATTTGCTGATCAGACAGAGCAAAAGCGTTGTTTGT	481		

Db	241	AGGAGCTACCATCTGAGCCCGCACCTTCTGACTCACAGAGAGAAAACCGCTTGTGT	300
Qy	482	GCAGCCACTGGCTGTGATCAAAAATTCACAACAAAATCAAACTTGAAAGAAACATTTTGAA	541
Db	301	GCAGCCAAATGGCTGTGATCAAAAATTCACAACAAAATCAAACTTGAAAGAAACATTTTGAA	360
Qy	542	CGCAAAACATGAAATTCACAACAAAACAAATATATATGCTAGTTTGAAGACTGTAAAGAAC	601
Db	361	CGCAAAACATGAAATTCACAACAAAACAAATATATATGCTAGTTTGAAGACTGTAAAGAAC	420
Qy	602	TTTAAAGAAACATCACAGCTGAAATTCATGCTCCAGCATACCAATGAACCTTATTC	661
Db	421	TTTAAAGAAACATCACAGCTGAAATTCATGCTCCAGCATACCAATGAACCTTATTC	480
Qy	662	AAGTGTACCCAGGAAGATGTGGGAACACTTGGATCACCCAGCAAGCTGAAACGACAT	721
Db	481	AAG-----	483
Qy	722	GCCAAAGCCCGCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATGG	781
Db	484	-----CTATGTTGTCAAAAAGATGTTCTTTGTGGCAAAAACATGG	525
Qy	782	ACGGAACCTCTGAACATGTGAGAGAAACCCCTAAAGAGGAATACTATGGAAGATATGC	841
Db	526	ACGGAACCTCTGAACATGTGAGAGAAACCCCTAAAGAGGAATACTATGGAAGATATGC	585
Qy	842	CGGAACACTTTTAAACGCAAAAGATTACCTTAAGCAACATGAAACATGATGCCAGAA	901
Db	586	CGGAACACTTTTAAACGCAAAAGATTACCTTAAGCAACATGAAACATGATGCCAGAA	645
Qy	902	AGGATATATGTGCTGCTGTCCCAAGAGAGGCTGTGAGAGAACTATACTGTGTTTAA	961
Db	646	AGGATATATGTGCTGCTGTCCCAAGAGAGGCTGTGAGAGAACTATACTGTGTTTAA	705
Qy	962	CTCCAAAGCATATCTCTCTCTCCATGAGAAACCCGCCCTTTGTGTGTAACATGCT	1021
Db	706	CTCCAAAGCATATCTCTCTCTCCATGAGAAACCCGCCCTTTGTGTGTAACATGCT	765
Qy	1022	GACTGTGGCAAAACATTTTGCAATGAAACAAAGTCTACAGGCTATGCTGTATCATGAT	1081
Db	766	GACTGTGGCAAAACATTTTGCAATGAAACAAAGTCTACAGGCTATGCTGTATCATGAT	825
Qy	1082	CCTGACAAAGAAATGAAGCTCAAAAGTCAAAAATCTCTGAAAAAAGCGAGTTTGCC	1141
Db	826	CCTGACAAAGAAATGAAGCTCAAAAGTCAAAAATCTCTGAAAAAAGCGAGTTTGCC	885
Qy	1142	TCTCATCTGATGATATATCTCTCCCAAGGAACCAAGGCAAGGCTTATCTTTGTCT	1201
Db	886	TCTCATCTGATGATATATCTCTCCCAAGGAACCAAGGCAAGGCTTATCTTTGTCT	945
Qy	1202	CAAAAAGGAGATCCCAACCTGTGTGGAAGCAAGATCTCTGCAAGTTGCACATCTT	1261
Db	946	CAAAAAGGAGATCCCAACCTGTGTGGAAGCAAGATCTCTGCAAGTTGCACATCTT	1005
Qy	1262	ACCCCTGGCTAA 1273	
Db	1006	ACCCCTGGCTAA 1017	
RESULT 5	AK057993	1920 bp mRNA linear PRI 31-OCT-2003	
LOCUS	AK057993	Homo sapiens cDNA FLJ25264 fis, clone STM05057, highly similar to	
DEFINITION	TRANSCRIPTION FACTOR IIIA.		
ACCESSION	AK057993		
VERSION	AK057993.1 GI:16553988		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens stomach mucosa cDNA to mRNA, clone_1lb:STM clone:STM05057.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1		

AUTHORS
 Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,
 Ono, Y., Houta, T., Hirooka, S., Murakawa, K., Takiguchi, S.,
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
 Yamashita, H., Chiba, Y., Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S.,
 Morikawa, M., Kawamura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Nishikawa, T., Sugiyama, A., Kawakami, B., Negai, K., Isogai, T., and
 Sugano, S.

TITLE
 NEDO human cDNA sequencing project

JOURNAL
 Unpublished

REFERENCE
 2 (bases 1 to 1920)

AUTHORS
 Sugano, S. and Suzuki, Y.

JOURNAL
 Direct Submission
 Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure, Human Genome
 Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail:cdna1@lelms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)

COMMENT
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB): cDNA library
 construction and 5'-end one pass sequencing: Institute of Medical
 Science, University of Tokyo, Laboratory of Genome Structure, Human
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for
 full insert sequencing: RAB and Helix Research Institute.

FEATURES
 Location/Qualifiers
 1..1920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="STM05057"
 /tissue_type="stomach mucosa"
 /clone_lib="STM"
 /note="cloning vector: pME18SFL3"

BASE COUNT
 520 a 462 c 438 g 500 t

ORIGIN

Query Match 69.98; Score 889.8; DB 9; Length 1920;
 Best Local Similarity 99.28; Pred. No. 5e-184;
 Matches 894; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

373 GGAGACGACATTGTTGTGACTGTGAAGGCTGTGGCAAGGCTTCATCGAGGACTACCA 432
 11 |||||||
 928 GGTAGACACATTGTTGTGACTGTGAAGGCTGTGGCAAGGCTTCATCGAGGACTGCA 987
 433 TCTGAGCGCCGACATCTTGACTGCACAGAGGAAAGCCGTTTGTGTGCGAGCCACTGG 492
 |||||||
 988 TCTGAGCGCCGACATCTTGACTGCACAGAGGAAAGCCGTTTGTGTGCGAGCCACTGG 1047
 493 CTGTGATCAAAAATTTCAACACAAAATCAAACTTGAAGAACTTTTGAAGCAAACTGA 552
 |||||||
 1048 CTGTGATCAAAAATTTCAACACAAAATCAAACTTGAAGAACTTTTGAAGCAAACTGA 1107
 553 AAATCAACAAAACATATATATGACGTTTGAAGACTGTGAAGACCTTTAAGAAACA 612
 1108 AAATCAACAAAACATATATATGACGTTTGAAGACTGTGAAGACCTTTAAGAAACA 1167
 613 TCACGACCTGAAAATTCATCAGTGCAGATACCAATGAACCTTATTCAAGTACCCA 672
 |||||||
 1168 TCACGACCTGAAAATTCATCAGTGCAGATACCAATGAACCTTATTCAAGTACCCA 1227
 673 GGAAGATGTGGGAAAACCTTTGCATCAGCAGACGAGAACTGAAGACATGCCAAGGCCA 732
 |||||||
 1228 GGGAGGATGTGGGAAAACCTTTGCATCAGCAGACGAGAACTGAAGACATGCCAAGGCCA 1287
 733 CGAGGCTATGTATGCAAAAAGGATGTCCTTTGGCAAAAACATGAGCGGAACCTTC 792
 |||||||
 1288 CGAGGCTATGTATGCAAAAAGGATGTCCTTTGGCAAAAACATGAGCGGAACCTTC 1347
 793 GAAACATGTGAGAGAAACCCATAAGAGAAATACTATGTGAAGTATCCGGAACATTT 852
 |||||||
 1348 GAAACATGTGAGAGAAACCCATAAGAGAAATACTATGTGAAGTATCCGGAACATTT 1407
 853 TAAACGAAAGATTACCTTAAGCAACATGAAAACCTATGCCCCAGAAAGGATGTATG 912
 |||||||

Db 1408 TAAACGAAAGATTACCTTAAGCAACATGAAAACACTATGCCCCAGAAAGGATGTATG 1467

Qy 913 TCGCTGTCCAAAGAGAGCTGTGAACACCTATACACTGTGTTTAAATCTCCAAAGCCA 972
 |||||||

Db 1468 TCGCTGTCCAAAGAGAGCTGTGAACACCTATACACTGTGTTTAAATCTCCAAAGCCA 1527

Qy 973 TATCTCTCTCTCCATGAGAAAGCCGCCCTTTGTGTGTGAACATGCTGGCTGTGGCAA 1032
 |||||||

Db 1528 TATCTCTCTCTCCATGAGAAAGCCGCCCTTTGTGTGTGAACATGCTGGCTGTGGCAA 1587

Qy 1033 AACATTTGCAATGAACAAAGCTCTACTAGGACATGCTGTGTATCATATCCCTGACAAGAA 1092
 |||||||

Db 1588 AACATTTGCAATGAACAAAGCTCTACTAGGACATGCTGTGTATCATATCCCTGACAAGAA 1647

Qy 1093 GAAATGAGCTGCAAGTCAAAAAAATCTCGTGAAGAAAGGAGTTGGCTCTCATCTCAG 1152
 |||||||

Db 1648 GAAATGAGCTGCAAGTCAAAAAAATCTCGTGAAGAAAGGAGTTGGCTCTCATCTCAG 1707

Qy 1153 TGTATATATCCCTCCCAAGAAAGGAGGCAAGGCTTATCTTTGTGCMAAAGAGAGA 1212
 |||||||

Db 1708 TGTATATATCCCTCCCAAGAAAGGAGGCAAGGCTTATCTTTGTGCMAAAGAGAGA 1767

Qy 1213 GTCAACCAACTGTGTGAAGCAAGATGCTCTGCACAGTTGCAGTACTTACCTTGGCTA 1272
 |||||||

Db 1768 GTCAACCAACTGTGTGAAGCAAGATGCTCTGCACAGTTGCAGTACTTACCTTGGCTA 1827

Qy 1273 A 1273

Db 1828 A 1828

RESULT 6
 AF391799 1298 bp mRNA linear ROD 31-JAN-2002

LOCUS
 AF391799

DEFINITION
 Mus musculus transcription factor I11A mRNA, partial cds.

ACCESSION
 AF391799

VERSION
 AF391799.1 GI:18448381

KEYWORDS
 Mus musculus.

SOURCE
 Mus musculus.

ORGANISM
 Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 1298)
 Hanas, J.S., Hocker, J.R., Cheng, Y.G., Lerner, M.R., Brackett, D.J.,
 Lightfoot, S.A., Hanas, R.J., Madhusudan, K.T. and Moreland, R.J.,
 cDNA cloning, DNA binding, and evolution of mammalian transcription
 factor I11A

TITLE
 factor I11A

JOURNAL
 Gene 282 (1-2), 43-52 (2002)

MEDLINE
 21673987

PUBMED
 11814676

REFERENCE
 2 (bases 1 to 1298)

AUTHORS
 Hanas, J.S., Hocker, J.R., Lerner, M.R., Brackett, D.J.,
 Lightfoot, S.A., Hanas, R.J., Madhusudan, K.T. and Moreland, R.J.,
 Direct Submission

TITLE
 Submitted (14-JUN-2001) Biochemistry/Molecular Biology, University
 of Oklahoma Health Science Center, 800 Research Parkway, Suite 448,
 Oklahoma City, OK 73104, USA

FEATURES
 Location/Qualifiers
 1..1298
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1..1203
 /note="Tf11A: Cys2His2 zinc finger protein; similar to
 xenopus laevis transcription factor I11A Swiss-Prot
 Accession Number P03001"
 /codon_start=1
 /product="transcription factor I11A"
 /protein_id="AA169686.1"
 /db_xref="GI:18448382"
 /translation="ERGTVEYQSIRAPVACVPPSRGRHVAAPROPGRDRALEPRSYVA
 EAVSLLTADAVGVAGCEGPAPPAIPSPFICSPDCSASYKAMKLDLHILKHGER
 PFVYDEGGKAFIRDYHLIRYHLITHTGKPFVCADDGCGNFKNTSNLKHIERHKG
 NPQKGYLCSEGGKAFKKRHQQLRTHQCHTSEPLRFRCTHEGCGKIFASPRLRHGK

BASE COUNT	342 a	368 c	358 g	230 t
ORIGIN	<p> VHEBYLQCKGCSFMGKWTWTELLKIMRCAHHEDITCNVCQRMFKRRDYLKQNHKTNAPE RDVYRCPRQCGCGRTYTTVFNLQSHILSFHEEKRFVEYHAGCGCTFMKOSLMHNSV HDPCKRMKGLKVRAPRRRSLASRLSGYPPKRRQEDYSLPNASSASSSSPEQLPP PAALLTVC" </p>			

Query Match	56.58;	Score 719.2;	DB 10;	Length 1298;
Best Local Similarity	79.58;	Pred. No. 9.3e-147;		
Matches 879;	Conservative 0;	Mismatches 218;	Indels 9;	Gaps 2;

OY	170	GGCCCCGGATCCGGCCGCGCGTGGCCAGATGGGTGTCTCTTGTACCATCGCCGAC	223
OY	170	GGCCCCGGATCCGGCCGCGCGTGGCCAGATGGGTGTCTCTTGTACCATCGCCGAC	223
Db	103	GTCCCTCTGGAGCCCGGGGTGTACGTGCGGGAAGCGGTCTGTGCTTACCATCGCGAT	162
OY	230	GCCTTTCATTCACACCGCGCGAGAGCTAGCTCCGACCCCGCGCCGCTTCCCGAGG	289
Db	163	GCCTTTCATTCACACCGCGCGAGAGCTAGCTCCGACCCCGCGCCGCTTCCCGAGG	289
OY	290	AGTTTCACTGTCTCTTCCCTGACTGACGAGCCCAATTACAGCAAAAGCTGAGTGC	349
Db	217	AGTTTCACTGTCTCTTCCCTGACTGACGAGCCCAATTACAGCAAAAGCTGAGTGC	349
OY	350	GGCAGACCTGTGCAAGCAGCAGGGGGAGAGCAATTTGTTGTGACATGAAAGGCTGGC	409
Db	277	GGCAGACCTGTGCAAGCAGCAGGGGGAGAGCAATTTGTTGTGACATGAAAGGCTGGC	409
OY	410	AAGGCTTCATCAGGAGCTACCATCTGAGCCGCGACATTTCTGACTACAGCAGAGAAAG	469
Db	337	AAGGCTTCATCAGAGAGCTACCATCTGAGCGGCGCATGTCTGATTACACCGGGGAAAG	396
OY	470	CCGTTTGTGTGTGACGCCACTGGCTGTGATCAAAATTTAACACAAATCAACTTTGAG	529
Db	397	CCGTTTGTGTGTGAGVATGATGGCTGTATCAAAATTTACACAAATCAAACTTTGAG	456
OY	530	AAATCTTTGAAGCAGCAACATGAAATTCAAAAAAACATATATATGCACTTTTAACAC	589
Db	457	AAACACATTTGACCGCAACACATGAAACCCACAAAAACAGTATTTTGTGACGTTATGAGGT	516
OY	590	TGTAGAAGACCTTTAAGAAACATCAGCAGCTGAAATTCATCACTGCGCAGCATACAT	649
Db	517	TGCAAGAAAGGCTTTAAGAGCAGCAGCAGAGCTGAGAAACCATCAGTGCACGACACAGC	576
OY	650	GAACCTTATTCAGGTACCACGAGAGATGTGGGAAACACTTTGTCAACCCGACAG	709
Db	577	GAGCGGCTTTCAGGTGTACCCAGAGGGATGTGGGAAACACTTTGTCCGCCAGCAGG	636
OY	710	CTGAAAGACATGCGCAAGGCCGCGAGAGGGTATGTATGTCAAAAAAGATGTTCCCTTGTG	769
Db	637	CTGAAAGCGCATGGGAAAGTTCAACAGAGGCTACCTGTGTCAAAAGSAGATGTTCTTTCATG	696
OY	770	GCAAAAACATGACGAGCACTTCTGAAACATGTGAGAAACCATTAAGAGAAATYACTA	829
Db	697	GGAAAAACGTGGAGAGAGTCTCTGAAACACATGAGGAAAGCCCATTAAGAGAGCAATACC	756
OY	830	TGTGAGATATCCGGAAGAACATTTAAACGCAAGATTAACCTTAAGACACATGAAACT	889
Db	757	TGTGATGTATGTACAGAGATGTTCAAGCGAGATTAACCTTAAGAGCAGCATGAAAGCT	816
OY	890	CATGCCCCAGAAAGGATGTATGTGCTGCTGTCCAAAGAGAGGCTGTGAGAAACATCTACT	949
Db	817	CACGCCCCGAAAGGATGTATGTACCGGTGTCCCGGCAAGGGCTGGGAAGAACCTACACA	876
OY	950	ACTGTGTTTATCTCCAAAGCCATATCTCTCTTCATGAGAGAAAGCGCCCTTTGTG	1009
Db	877	ACCGGTTCACCTCGCAGAGCCACCATTTCTCTCTTCACAGAGAAAGCGCCATTTGTG	936
OY	1010	TGTGAAACATGCTGTGCTGTGGCAAAACATTTGCATGAACAAGCTCATGTAGCATCT	1069
Db	937	TGTGAGCAGCTGTGCTGTGGCAGAGCACTTGTGCAATGAACAGAGTGTCTCATGTAGCAGCACT	996
OY	1070	GTTGTACATGATCTCTGACCAAGAAAGAAATGAGCTCAAAAGTCAAAAATCTCTGTGAAAA	1129

Dd	997	GTCGTGCACCATCCCCGACAAGAAAGAGATGAAGCTCAAAAGTAAGAAGCCCTTCGGAGAGA	1056
Qy	1130	CGAGATTTTGGCCTCTCATCTCACTGGATFATATTCCTCCCAAAAAGGAAACAAGGGCAAGC	1189
Dd	1057	CGCAGCTTGGGCTCTGCCCTCAGTGGGTACTTCCCTCTAAGAGAAACAAGAGCCGAC	1116
Qy	1190	TTATCTTTGTGTCA---AAAGGAGAGTACCACCACTGTGTGGAAGACAAGATGCTTCG	1246
Dd	1117	TACTCTTGCTAACGCGACGCGAGATCCAGAGAGAGCCGAGAGGCGACGCTGGCCCG	1176
Qy	1247	ACAGTTGCAGTACTTACCTTTGGCTA	1272
Dd	1177	CCAGCCGCTTACTACACTGTCTCTA	1202

RESULT 7					
AF391798					
LOCUS	AF391798	1071 bp	mRNA	linear	ROD 31-JAN-2002
DEFINITION	Rattus norvegicus transcription factor IIA mRNA, partial cds.				
ACCESSION	AF391798				
VERSION	AF391798.1	GI:18448379			
KEYWORDS					
SOURCE	Rattus norvegicus.				

ORGANISM *Rattus norvegicus*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1071)
Hanas,J.S., Hocker,J.R., Cheng,Y.G., Ierner,M.R., Brackett,D.J.,
Lightfoot,S.A., Hanas,R.J., Madhusudan,K.T., and Moreland,R.J.:
cDNA cloning, DNA binding, and evolution of mammalian transcription
factor IIA
JOURNAL Gene 282 (1-2), 43-52 (2002)

MEDLINE	21673987
PUBMED	11814676
REFERENCE	2 (bases 1 to 1071)
AUTHORS	Hanas,J.S., Hocker,J.R., Lerner,M.R., Brackett,D.J., Lightfoot,S.A., Hanas,R.J., Madhusudan,K. and Morland,R.J.

TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Biochemistry/Molecular Biology, University of Oklahoma Health Science Center, 800 Research Parkway, Suite 448, Oklahoma City, OK 73104, USA
FEATURES location/Qualifiers
source 1. .1071

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
<1..1011

```

```

/note="TFIIIA: Cys2His2 zinc finger protein; similar to
Xenopus laevis transcription factor IIA Swiss-Prot
Accession Number P03001"
/codon_start=1
/product="transcription factor IIA"
/protein_id="AAL6965.1"
/db_xref="GI:18448380"
/translation="AARGALPSRPTCSPPDCSASYNKAMKLDHLCKHTGERSVFCCI
VEGGKARIDRYHLNRHLITHTCEKRPFCANCGCNQKSTSYNLKHTIERHNENPKQK
IVCNESGCKAKFKKHQQLRTHQCOQHTNPLFRCITGEGCKHFPASPSRLTKRGKAVHEG
LCQKSGCVGKTTWELLKHTREAKHEEYTCVCKMFKRDLKOHMTAPADEVDYFV
CPRECAKRTYTVTFNLOSHILSFHDEKRPFCVAGCGCKTPAMKOSLIRHSVHVDPPDK
KRMKLKLVPPRRERSLASRLSGYVPPKCKOEPDCLSPNSTLESSSPKATMLAPALLT
VH"

```

	Query Match	54.3%	Score 691	DB 10	Length 1071
	Best Local Similarity	80.9%	Pred. No.	1.3e-140	
	Matches	805	Conservative	0	Mismatches 190; Indels 0; Gaps 0;
QY	278	GGCGTTCACGAGGATTATCTGTCCTCTCCCTGACTGCGAGCGCCATTAACGAAAGCC	337		
DB	16	GGCGTTCACGAGGATTATCTGTCCTCTCCCTGACTGCGAGCGGTATGTTAAACAAAGCC	75		

```

Oy 338 TGGAGCTTGAGCCCGACCTGTGCAAGCACACGCGGGGAGAGACATTGTTGTGACTAT 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76 TGGAGCTAGACCGCCACCTGTGTAAGCACACGCGGGGAGAGAGTCTGTTGTGACTAT 135
Oy 338 GAAGGTGTGGCAAGCCCTTCATCAGGGACTTCCATCTGAGCCGCCACATTTGACTAC 457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 GAGGGCTGTGGCAAGCCCTTCATCAGGGACTTCCATCTGAGCCGCCACATTTGACTAC 195
Oy 458 ACAGAGAAAAAGCCGTTGTTGTGTCAGCAGCTGCTGTGATCAAAATTCGAACCAAA 517
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 ACCGAGAAAAAGCCATTTGTTGTTGTGTCAGATATATGCTGTATCAAAATTCGACCAAA 255
Oy 518 TCAACTTGAAGAAACATTTTGAACGCAAAATGAAATTCACAAAACATATATATGC 577
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 TCAACTTGAAGAAACATTTTGAACGCAAAATGAAATTCACAAAACATATATATGC 315
Oy 578 AGTTTGAAGACTGTAAAGAACCTTTAAAGAACATCAGCAGCTCAAAATTCATGATGC 637
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 AATTTTGAAGGTTCAGAAAGCCCTTTAAAGAACACACAGCAGCTAGAAACCATCATGTC 375
Oy 638 CAGCATACCAATGAACTCTATTCAAGTGTACCGAAGATGTTGGGAAACACTTTCGA 697
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 376 CAGCAGACAAATGAGCCACTCTTCAGGTGTACCCAGAGGATGTGGAGACACTTTGCC 435
Oy 698 TCACCCAGCAAGCTGAAGCAGACATCCCAAGCCCAAGAGGCTATGTATGTCAAAAGGA 757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 TCACCCAGCAAGCTGAAGCAGACATCCCAAGGCTGTACAGAGGCTACTATGTCAAAAGGA 495
Oy 758 TGTCTCTTTGTGGCAAAACATGAGCAGCACTTCTGAAACATGTAGAGAAACCATAAA 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 TGTCTCTTTGTGGCAAAACATGAGCAGCACTTCTGAAAGCAGCAGAGAACCCCATAAA 555
Oy 818 GAGAAATATCTATGTGAAGTATGCGGAAACATTTAAACGCAAGATTACTTAAAGCA 877
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 GAGAGAGTGAAGCTGACAGAGTGTCCAGAAAGATGTTCAAGCCGAAAGATCAGCTTAAAGAG 615
Oy 878 CACATGAAGAACTGATGCCCCAGAAAAGGATGTATGCTGTCAAGAGAGCGTGTGA 937
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 616 CACATGAAGAACTGATGCCCCAGGAGGATGTATGCTGTCCGCGGGAAGCGTGTGA 675
Oy 938 AGAAGCTATGCTATGCTTTTATCTCCAAAGCCATCTCTCTCTCCATGAGAGAAAGC 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 AGGACCTTACACCGAGCTGTTCACCTCGAGAGCCACATTTCTCTCTCCACAGGAGAAAG 735
Oy 998 CGCCCTTTGTGTGAAACATGCTGGCTGTGGCAAAACATTTGCAATGAACAAAGCTGC 1057
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 CGCCCATTTGTGTGAGAGTGTGCTGTGTGGCAAGACCTTTGCATGAACAAAGTCTC 795
Oy 1058 ACTAGCATGCTGTTGTACATGATCTTGACAAAGAAAGAAATGAAGCTCAAAATAAA 1117
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 ATGAGGACAGATGCTGCTGACAGATCTTGACAAAGAGAGATGAAGCTCAAGATTAGACCA 855
Oy 1118 TGTGCTGAAAAAGGAGTTTGGCCCTCTCATCTGATGATATATCCCTCCCAAAAGGAAA 1177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 CTTCTGGGAGAGAGCGAGCTGTGGCTGTGGCCCTGACCGGCTGACCTCTTCAAAAGGCAAA 915
Oy 1178 CAAGGCAAGAGCTTATCTTTGTGTCAAAAGGAGAGTCAACCAACTGTGTGGAAGACAAG 1237
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 CAAGAGCCCGAGTCTGCTCTTGTGCTTAACAGACACAGAGTCCAGAGACAGCCCAAGGCCACG 975
Oy 1238 ATGCTCTGACAGTTCAGATGATTAACCTTGCGCTA 1272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 976 ATGCTCGCCCGAGCTGCGTACTACTGTGCCACTA 1010
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 8
BC032292          996 bp  mRNA  linear  ROD 07-AUG-2002
LOCUS             Mus musculus, general transcription factor III A, clone MGC:40923
DEFINITION        IMAGE:5374268, mRNA, complete cds.
ACCESSION         BC032292
VERSION           BC032292.1  GI:21595507

```

KEYWORDS
MGC.
house mouse.
Mus musculus.

SOURCE
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 996)

REFERENCE
AUTHORS
Strausberg, R.
TITLE
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nih.gov
Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dieterich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karins, E., Kwong, P., Latic, P., Legaspi, R.,
Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stancir, P., Thomas, P.J., Touchman, J.W.,
Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 65 Row: 1 Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES

source
1..996
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N-3"
/clone="MGC:40923 IMAGE:5374268"
/tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_1lb="NCI_CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
365..874
/codon_start=1
/product="general transcription factor III A"
/protein_id="AAH32292.1"
/db_xref="GI:21595508"
/translation="MGRTWTELLKHMREAHKEDIYCNVQRMFKRBDYLKOHKTHAP
ERDYVRCRPGCGRTYTVFNLOSHTLSFHEBRPFCVCEHAGGAKTPAKMSIRSY
VHPDPRKMLKXRAPREPRERSLRSLSGYPPPRKQPDPSLPLNASESSSPEDQLP
PRATLIVC"

CDS
365..874
/codon_start=1
/product="general transcription factor III A"
/protein_id="AAH32292.1"
/db_xref="GI:21595508"
/translation="MGRTWTELLKHMREAHKEDIYCNVQRMFKRBDYLKOHKTHAP
ERDYVRCRPGCGRTYTVFNLOSHTLSFHEBRPFCVCEHAGGAKTPAKMSIRSY
VHPDPRKMLKXRAPREPRERSLRSLSGYPPPRKQPDPSLPLNASESSSPEDQLP
PRATLIVC"

BASE COUNT
319 a 261 c 244 g 172 t
ORIGIN

Query Match 44.0% Score 560.6; DB 10: Length 996;
Best Local Similarity 78.9%; Pred. No. 4.2e-112;
Matches 681; Conservative 0; Mismatches 179; Indels 3; Gaps 1;
Oy 413 GCCTTCATCAGGACTACATCTGAGCCGCCACATCTGACTCAGACAGAGAAAGCCG 472
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11 GCCTTCATCAGGACTACATCTGAGCCGCCACATCTGACTCAGACAGCGGAAAGCCG 70
Oy 473 TTTGTTTGTGACGCACTGGCTGTGATCAAAATTCACACAAATCAACTTGAAGAAA 532
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71 TTTGTTTGTGACGATGATGCTGTAATCAGAAATTCACACAAATCAACTTGAAGAAA 130
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||


```
repeat_region      8718..9750      /rpt_family="L1"
repeat_region      9751..12094     /rpt_family="L1"
repeat_region      12095..12372     /rpt_family="Alu"
repeat_region      12373..12436     /rpt_family="CA)n"
repeat_region      12437..12842     /rpt_family="L1"
repeat_region      12888..13001     /rpt_family="L1"
repeat_region      13003..13358     /rpt_family="MaLR"
repeat_region      13365..14214     /rpt_family="L1"
repeat_region      14264..14417     /rpt_family="L1"
repeat_region      14415..14729     /rpt_family="L1"
repeat_region      14816..15026     /rpt_family="MERL_type"
repeat_region      15648..15729     /rpt_family="CA)n"
repeat_region      16041..16362     /rpt_family="L1"
repeat_region      16679..16732     /rpt_family="L1"
repeat_region      17025..17315     /rpt_family="Alu"
repeat_region      17465..17685     /rpt_family="L1"
repeat_region      18044..18633     /rpt_family="L1"
repeat_region      18630..18750     /rpt_family="L1"
repeat_region      18760..19497     /rpt_family="L1"
repeat_region      19508..19692     /rpt_family="L1"
repeat_region      19702..19814     /rpt_family="L1"
repeat_region      21491..21543     /rpt_family="L2"
repeat_region      21987..24424     /rpt_family="L1"
repeat_region      24425..24720     /rpt_family="Alu"
repeat_region      24722..24821     /rpt_family="L1"
repeat_region      25383..25745     /rpt_family="L1"
repeat_region      29202..29272     /rpt_family="MERL_type"
repeat_region      30338..30386     /rpt_family="CA)n"
repeat_region      30984..31045     /rpt_family="MERL_type"
repeat_region      31041..31398     /rpt_family="MERL_type"
repeat_region      32178..32319     /rpt_family="MERL_type"
repeat_region      32475..32515     /rpt_family="MIR"
repeat_region      32589..32646     /rpt_family="L2"
repeat_region      32940..33224     /rpt_family="L2"
repeat_region      33475..33558     /rpt_family="Alu"
misc_feature       33843..33929     /note="match to EST AA928509 (NTD:g3076800) om17907.sl"
repeat_region      35891..36381     /rpt_family="MERL_type"
```

```
repeat_region      37910..38123     /rpt_family="L2"
repeat_region      38302..38768     /rpt_family="MERL_type"
repeat_region      42711..43026     /rpt_family="L1"
repeat_region      43023..43418     /rpt_family="L1"
repeat_region      43552..43866     /rpt_family="L1"
repeat_region      43847..44539     /rpt_family="L1"
repeat_region      44534..45153     /rpt_family="L1"
repeat_region      45132..45245     /rpt_family="L1"
repeat_region      45246..45667     /rpt_family="L1"
repeat_region      45674..45869     /rpt_family="MaLR"
repeat_region      45871..46089     /rpt_family="L1"
repeat_region      46091..47111     /rpt_family="Alu"
repeat_region      47186..47354     /rpt_family="L1"
repeat_region      47529..47577     /rpt_family="MERL_type"
repeat_region      47673..48037     /rpt_family="MERL_type"
repeat_region      48067..48494     /rpt_family="L2"
repeat_region      51739..51768     /rpt_family="L2"
repeat_region      52362..52665     /rpt_family="CA)n"
repeat_region      /rpt_family="L1"
```

Query Match 36.9% Score 470.2: DB 9: Length 97979;
Best Local Similarity 76.8% Pred. No. 4.9e-92;
Matches 655; Conservative 0; Mismatches 158; Indels 40; Gaps 5;

```
QY 389 TGTGACTATGAAAGGTGTGGCAGGCCCTTCATCAGGACTACCATCTGAGCCGCACATT 448
      ||||||| ||||||| || ||||||| |||| ||||| |||||||
DB 79437 TGTGACTATTAAGGCTGTGCAAGCCTTCATCAGAGCTACCTCTGATGCCACATC 79378
QY 449 CTGACTCACACAGAGAAAGCCCTTGTGTGCGAGCCACTGGCTGTGATCAAA----- 503
      ||| |||| ||||||| || ||||||| || ||||||| |||
DB 79377 CTGATTCACATTGGAGAAAGCCCTTTATTTGTGCGAGCTAGTGGCTGTGATAAAAA 79318
QY 504 -----AATTCAACAAATAATCAAACTTGAAAGAAACATTTGAAAGCAATGAAATC 557
      ||| ||| ||||| ||| ||||||| ||||| ||||||| ||||| |||||||
DB 79317 AAAAAGAAATCCACAGAAATCCAACTGGAGAAACATTTGCATCAACATGAAATC 79258
QY 558 AACAAAAACATATATATGACGTTTTGAAGACTGTAAAGACCTTTAAGAAACATCAGC 617
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79257 TGCAAAAGCAATATGTATGCAATTTTGAAGGATGTAAAGACCTTTAAGAACATCAGC 79198
QY 618 AGCTGAAATCCATCAGTGCAGCATACCAATGAACCTCTATTCAAGTACCAGGAAG 677
      ||||||| ||||||| || ||||| ||||||| || |||||
DB 79197 AGCTGAAATCCATCAGTGCAGCATGAAATTAACCCCTATTCAAGTTAACCTTGAAG 79138
QY 678 GATGTGGAAGACCTTGCATCCAGCCAGAGTGAAGAGCATGTGCCAAGGCCACAGAG 737
      ||||||| ||||||| ||||| ||||| ||||| ||||| |||||
DB 79137 GATGTGGAAGACCTTGTTCACCCCTCCAGGCTGGAGAGCAGATGCAGGTTCAACA 79078
QY 738 GCTATGTATGTCAAAAGAGATGTTCTTTGTGCAAAACATGAGCAAGCACTTGTGAAC 797
      | ||||||| ||||||| ||||||| ||||||| |||||||
DB 79077 AATTTATATGTCAATATAGGAGTGTCTTTGTGCAAAACATGAC----- 79032
QY 798 ATGTGAGAGAAACCATTAAGAGGAATATCTATGTGAAGTATGCCGAAACATTTAAAC 857
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79031 ---AGAGAGAAATCCATAGAAAGAAATA--ACNTGATGTATGCCGAAACATTTCAAT 78977
```

```

QY      858  GCAAGATTAACCAACACATGAAACCTCATGCCCGAAGAGGATGTATGCGCT 917
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78976  GTGAGATTATCTCAGATTAACCTGTGAAAACTCATGCCCTAGAAAGAGATGTATGTAAC 78917

QY      918  GTCCAAAGAGAGGCTGTGAGAAACCTATACCTGTGTAACTCCAAAGCCATATCC 977
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78916  ATCCAAAGAGAGAGGCTGTGAGAAACCTATACCTGTGTAACTCCAAAGCCATATCC 78866

QY      978  TCTCCTTCATGAGGAAAGCCGCTTTGTGTGTGAACATGCTGGCTGTGGCAAAACAT 1037
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78865  TTTCTTTCTTAAGGAGAAACACCCATTTATGTGTGAACATGCTGGCTGTGGCAAAAGGT 78806

QY      1038  TTCAATGAAACAAAGCTCCTACTAGCATCTGTGTACATCTGCATCCAGAGAGAAAA 1097
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78805  TTGGTATGAAACAAACCTCTCACAAGACATGCTATGTGTATGAGCTTGACAGAGAGAAAA 78746

QY      1098  TGAAGCTCAAAAGTCAAAAATCTCTGTGAAAAACGAGCTTGGCCTCTCATCTCAGTGGAT 1157
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78745  TAAAGATCAAAAGTCAAAAATCTCTGTGAAAAATGAAATTTGGCCTCTCATCTCAGT-AAA 78687

QY      1158  ATATCCCTCCCAAAAGGAAACAGGCAAGGCTTATCTTTGTGTCAAAACGAGAGTCAAC 1217
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78686  ATATCCCTCTTAAGAGGAAACAGAGAGGCTTATCTTTGTGTCAATTTGAAATGAGAGATCAT 78627

QY      1218  CCAACTGTGTGA 1230
          | | | | | | | |
Db 78626  TGAATTGTATTGA 78614

RESULT 10
AC006045 193126 bp DNA linear PRI 05-MAY-1999
LOCUS     Homo sapiens clone NH0547615, complete sequence.
DEFINITION
AC006045 AC006045 GI:4753227
VERSION   AC006045.2 GI:4753227
KEYWORDS  HTG.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
AUTHORS   Waterston,R.H.
TITLE     2 (bases 1 to 193126)
REFERENCE JOURNAL Direct Submission
           Submitted (22-NOV-1998) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE JOURNAL Direct Submission
           Submitted (05-MAY-1999) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
COMMENT   On May 5, 1999 this sequence version replaced gi:3907455.
FEATURES
SOURCE    Location/Qualifiers
           1..193126
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="NH0547615"

BASP COUNT 63568 a 34097 c 33661 g 61800 t
ORIGIN
Query Match 36.9% Score 470.2; DB 9; Length 193126;
Best Local Similarity 76.8% Pred. No. 5,4e-92;
Matches 655; Conservative 0; Mismatches 158; Indels 40; Gaps 5;

```

```

QY      449  CTGACTCAGCAGAGAAAAAGCCGTTTGTGTGCAAGCCACCTGGCTGTATCAAA----- 503
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128614 CTGATTCACATTTGAGAGAAAAAGCCGTTTATTTGTGAGCTAGTGGCTGTGTAATAAAAAA 128755

QY      504  -----AATTCACACAAAAATTCAAACTGTGAAGAAACATTTTGAACCAACATGAAATTC 557
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128754 AAAAAGAAATCCCAAGAAATTCACACTGAGAGAAACATTTTGCATGCAAAACATGAAATTC 128695

QY      558  AACAAAAACAATATATATGCAAGTTTGAAGACGTGAAGAAAGACCTTTAAGAAACATCAGC 617
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128694 TGCAAAAAGCAATATGATGCAAAATTTGAAGAGATGAAGAGACCTTTAAGAAACATCAGC 128635

QY      618  AGGTGAAAAATCCACAGTGGCAGCATACCAATACCACTGATATCAAGTGAAGCCAGAGAG 677
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128634 AGCTGAAAAATCCACAGTGGCAGCATGAAATATTAACCCCTATTTCAAGTTTAACCTTGAAG 128575

QY      678  GATGTGGAAAACACTTTGTCATCACCCAGCAAGCTGAAGACATGCCAAGGCCAGAGAG 737
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128574 GATGTGGAAAACACTTTGTTCAACCCCTCCAGGCTGGGAGAGACAGATGCAAGCTTCATCAGA 128515

QY      738  GCTATGATGTCAAAAAGATGTTCTTTGTGCAAAAACATGAGAGGAACTTCTGAAC 797
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128514 AAATTAATGTCATATAGGAGATGTTCTTTGTGCAAGAACATGAGAC----- 128469

QY      798  ATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGATATCCGGAACATTTAAAC 857
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128468 ---AGAGAGAAATCCATAGAGAGAAATTA--ACATGATGTATGTCAGAAAAACATTCAAAT 128414

QY      858  GCAAGATTACTTTAAGCACAACATGAAACCTCATGCCCGAAGAGGATGTATGCGCT 917
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128413 GTGAGATTATCTCAGATTAACCTGTGAAAACTCATGCCCTAGAAAGAGATGTATGTCAC 128354

QY      918  GTCCAAAGAGAGGCTGTGAGAAACCTATACCTGTGTAAATCTCCAAAGCCATATCC 977
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128353 ATCCAAAGAGAGAGCTGAGAGAA-----ACTGTATTTAATCTCCAAAGCTGTCTTC 128303

QY      978  TCTCCTTCATGAGGAAAGCCGCTTTGTGTGTGAACATGCTGGCTGTGGCAAAACAT 1037
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128302 TTTCTTTCTTAAGAGGAAAGCCACCTATTATGTGTGAACATGCTGGCTGTGGCAAAAGGT 128243

QY      1038  TTCAATGAAACAAAGCTCCTACTAGCATCTGTGTATCATGATCTGCACAAAGAGAAAA 1097
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128242 TTGGTATGAAACAAACCTCTCACAAGACATGCTATGTGTATGAGCTTGACAGAGAGAAAA 128183

QY      1098  TGAAGCTCAAAAGTCAAAAATCTCTGTGAAAAACGAGCTTGGCCTCTCATCTCAGTGGAT 1157
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128182 TAAAGATCAAAAGTCAAAAATCTCTGTGAAAAATGAAATTTGGCCTCTCATCTCAGT-AAA 128124

QY      1158  ATATCCCTCCCAAAAGGAAACAGGCAAGGCTTATCTTTGTGTCAAAACGAGAGCTCAC 1217
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128123 ATATCCCTCTTAAGAGGAAACAAAGAGAGGATTTATCTTTGCATGTGAAGAGAGATCAT 128064

QY      1218  CCAACTGTGTGA 1230
          | | | | | | | |
Db 128063 TGAATTGTATTGA 128051

RESULT 11
AC074390 111722 bp DNA linear PRI 21-FEB-2002
LOCUS     Homo sapiens BMC clone R11-51SK14 from 7, complete sequence.
DEFINITION
AC074390 AC074390 GI:18250153
VERSION   AC074390.6 GI:18250153
KEYWORDS  HTG.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)

```

MEDLINE 09063792
 PUBMED 9647074
 REFERENCE 2 (bases 1 to 111722)
 AUTHORS Bielicki,L., Haekenson,W. and Boyer,E.
 TITLE The sequence of Homo sapiens BAC clone RP11-515K14
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 111722)
 AUTHORS Waterston,R.H.
 JOURNAL Direct Submision
 Submitted (30-JUL-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 111722)
 AUTHORS Waterston,R.H.
 JOURNAL Direct Submision
 Submitted (19-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 111722)
 AUTHORS Waterston,R.
 JOURNAL Direct Submision
 Submitted (21-FEB-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jan 19, 2002 this sequence version replaced gi:17976493.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0515K14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is CTD-2105K18, 2000 bp overlap; the clone sequenced to the right is RP11-533K11, 2000 bp overlap. Actual start of this clone is at base position 141753 of CTD-2105K18; actual end is at base position 16899 of RP11-533K11. Polymorphisms have been identified between AC016444 and AC074390.

FEATURES Data from AC016444 was used to finish this clone, AC074390.
 source Location/Qualifiers
 1. 111722
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone="RP11-515K14"
 /clone_lib="RPCI-11"
 1. 249
 repeat_region
 /rpt_family="MaLR"
 250. 640
 /rpt_family="MaLR"
 1650. 2014
 /rpt_family="ERV1"
 2015. 3233
 /rpt_family="L1"
 3234. 3578
 /rpt_family="ERV1"
 3586. 3835
 /rpt_family="ERV1"
 4560. 4873
 /rpt_family="ERV1"
 4880. 4975
 /rpt_family="GA-rich"
 4991. 5133
 /rpt_family="ERV1"
 5577. 5801
 /rpt_family="L1"
 6687. 6982
 /rpt_family="Alu"
 7392. 7449
 /rpt_family="AT-rich"
 7495. 7564
 /rpt_family="AT-rich"
 7559. 8728
 /rpt_family="L1"
 8746. 9038
 /rpt_family="Alu"
 9017. 9053
 /rpt_family="(TAA)n"
 9061. 9362
 /rpt_family="Alu"
 9372. 10143
 /rpt_family="L1"
 9934. 9963
 /rpt_family="(TC)n"
 10144. 10454
 /rpt_family="Alu"
 10427. 10454
 /rpt_family="(A)n"
 10455. 11140
 /rpt_family="L1"
 11487. 11960
 /rpt_family="L2"
 12112. 12305
 /rpt_family="L2"
 12306. 12615
 /rpt_family="Alu"
 12591. 12615
 /rpt_family="(A)n"
 12616. 12766
 /rpt_family="L2"
 13024. 13088
 /rpt_family="(TA)n"
 13613. 13641
 /rpt_family="AT-rich"
 14281. 14328
 /rpt_family="(TA)n"
 14329. 15028
 /rpt_family="MER2_type"
 15050. 15494
 /rpt_family="MER2_type"

	repeat_region	15445..15464	/rpt_family="(T)n"
	repeat_region	13495..16698	/rpt_family="L1"
	repeat_region	17544..17927	/rpt_family="L1"
	repeat_region	19584..19618	/rpt_family="L1"
	repeat_region	21646..21769	/rpt_family="MIR"
	repeat_region	21860..22043	/rpt_family="MERI_type"
	repeat_region	22044..22427	/rpt_family="MALR"
	repeat_region	22428..22467	/rpt_family="MALR"
	repeat_region	23300..23664	/rpt_family="MALR"
	repeat_region	23870..23910	/rpt_family="(TA)A)n"
	repeat_region	24405..24635	/rpt_family="L1"
	repeat_region	24636..24901	/rpt_family="MALR"
	repeat_region	25481..25504	/rpt_family="AT_rich"
	repeat_region	26543..26672	/rpt_family="MERI_type"
	repeat_region	26769..26911	/rpt_family="CRL"
	repeat_region	26928..27295	/rpt_family="MALR"
	repeat_region	27653..27678	/rpt_family="AT_rich"
	repeat_region	27934..28093	/rpt_family="MERI_type"
	repeat_region	29167..29194	/rpt_family="AT_rich"
	repeat_region	29191..29309	/rpt_family="ALU"
	repeat_region	29348..29644	
Query Match		35.6%	Score 452.6; DB 9; Length 111722;
Best Local Similarity		80.3%	Pred. No. 3.5e-88;
Matches 685; Conservative		0; Mismatches 139; Indels 29; Gaps 12;	
QY	269	CGGCGCCGGGCTTCGAGGAGGTTCATCTCTCTCCCTGCAGTGCACGCCCAATTAC	328
DB	87601	COCCTTCACACTCTTTAGGAGGGCTCTCTCTCTCC-CAGTGTGACCACTTAG	87543
QY	329	AGCAAAGCCTGGAAAGCTTGACCGCACCTGTGTCAGACACACGCGGAGAGACCATTGTT	388
DB	87542	AAAAAGCCTGGAAAGCTTAATGTGCACCTGT--AATCATGTGGGAAGAAGAACCTTTGTT	87485
QY	389	TGTGACTATGAAGGTTGTGGCAAGGCTTCATC-AAGGACTACATCTGAGCGCCACAT	447
DB	87484	TATGACTGTGAAGGTTGTGGCAAGGACCACTCAAGAGACTAACATCTGAGTTCATGT	87425
QY	448	TCTGACTACACAGAGAAAACCCCTTTGTTGTGTCAGGCACTGGCTGTGATC-AAAAAT	506
DB	87424	CCTGA-TCTACTGTGAGACAAACCCCTATGTTGTGTCAGCTAATCTGTGATCTAAAAAT	87366
QY	507	TCAACACAAAATCAACAATTTGAAGAAATTTTGAACGCAAAACATGMAAATCAACAAAAAC	566
DB	87365	TCAACACAAAATCAACAATTTGAAGAAACATTTTACAGCGCAAAACATGMAAACCAAAAAA	87306
QY	567	A-----ATAATATGTCAGTTTGAAGACTGTAGAAGACCTTTAAGAAAACATCAGCAGCT	621
DB	87305	AAAAATATACATATATCAATTTTGAAGGTTGAAGAAGACCTTTAGC-AACATCAGCAGCT	87247
QY	622	GAAATTCATATGTCAGTCCACGACATACCAATGAACCTTATTTCAAGGTATCCAGGAAGATG	681
DB	87246	GAAATTCATATGTCAGTCCACGACATACCAATGAACCTTATTTCAAGGTATCCAGGAAGATG	87187

QY	682	TGGGAAACACTTTCATGATACCAGCAAGCTGAAACGACATGCGCAAGGCCGAGAGGCTA	741
Db	87186	TGGGAAGCATTTTTCCTCCATCCATCCACG-----AAGCGGCAAGGCCATTCGGGGCTC	87134
QY	742	TGTATGTCAAAAAGGATGTCCTTCTGTGGCAAAAACATGACGGAACTCTGAACATGT	801
Db	87133	TATATGTCAAAAAGCGTGTCTTCTTGTGGTGAAGACATGACAGAAAGTTTGAACATGT	87074
QY	802	GAGAGAAACCCCTAAGAAGGAATACTGTGTGAATATGCCGGAAACATTTAAACGCA	861
Db	87073	GAGAGAAACTTATATAAAGAAACCAACTATGAACAGCGCCAGAAAACATTTAAAC----	87018
QY	862	AGATTACCTTAAAGCAACACATGAAACATCATGCCCAAGAAAGGATGTATGCTGCTCC	921
Db	87017	--ATTATCTTGAAGCAACATGTGAACACTCATGCCCAAGAAAGGATAT-TTTTGACATAC	86961
QY	922	AAGAGAGGCGTGTGGAGAACCCTATACTACTGTGTTTAAATCTCCAAAGCCATATCTCTC	981
Db	86960	AAGAGAGCTCTCTGGAAACACCTACCAACACTGTCTTACTCTACAAATCCATATCTCTTC	86901
QY	982	CTTCCATGTGGAAACCCGCCCTTTGTGTGTGCAACATGCTGGCTGTGGCAAAACATTGC	1041
Db	86900	CAT--GAGAAAAAAGGTGCTATTTGCATGTGAAACATGTGGCTGTGGCAAAACATTGC	86843
QY	1042	AATGAACAACACTCTCACTAGGCATGCTGTGTATACATCTCTGACAGAAGAAATGAA	1101
Db	86842	AATGAACAACAACTCTTCTTAGAGCATGTGTGTACAT-ATTCTGACAGAAGAAATGAA	86784
QY	1102	GCTCAAAAGTCAAA 1114	
Db	86783	GCTCAGAGTAATA 86771	

```

RESULT 12
AC016444      206622 bp   DNA    linear     HTG_04-MAY-2001
LOCUS         Homo sapiens chromosome 11 clone RP11-384I14 map 11, WORKING DRAFT DEFINITION
AC016444      AC016444          Homo sapiens chromosome 11 clone RP11-384I14 map 11, WORKING DRAFT ACCESSION
AC016444      AC016444          SEQUENCE, 16 unordered pieces.
VERSION       AC016444.3 GI:10045722
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarchonta; Primates; Carnivora; Hominoidea; Hominidae; Homo.

REFERENCE
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Balaban,R., Barlow,A., Beckwith,T.D., Berg,P., Binkley,S., Boulton,J.R., Brown,W.A., Casper,A.M., Cawthon,G., Cohen,D.L., Collins,F.S., Collumore,A., Cook,K.P., DeBry,J.W., Dewar,K., Dieckhaus,U., Dolan,P.J., Doyle,M.K., Fellers,V., FitzHugh,W., Fontana,W., Forrest,C., Funkhouser,R., Gage,D., Galante,J.C., Gardy,J.E., Grady,S., Grant,G., Hagos,B., Heathcote,A., Horton,L., Howell,I., Johnson,R., Jones,C., Kahn,L., Karasik,A., Klein,J., Lechoczy,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.N., McElwain,P., McKusick,A., McKernan,K., McLoughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testafaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.U., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (26-Nov-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6778507.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5006
Center clone name: 384_1-14

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190829 bases at least Q40
Consensus quality: 198597 bases at least Q30
Consensus quality: 202013 bases at least Q20
Insert size: 213000; agarose-fp
Insert size: 205122; sum-of-ctrls
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 13492: contig of 13492 bp in length
* 13493 13592: gap of 100 bp
* 13593 14758: contig of 1166 bp in length
* 14759 14858: gap of 100 bp
* 14859 16187: contig of 1329 bp in length
* 16188 16287: gap of 100 bp
* 16288 18628: contig of 2341 bp in length
* 18629 18728: gap of 100 bp
* 18729 22545: contig of 3817 bp in length
* 22546 22645: gap of 100 bp
* 22646 27333: contig of 4688 bp in length
* 27334 27433: gap of 100 bp
* 27434 32718: contig of 5285 bp in length
* 32719 32818: gap of 100 bp
* 32819 37975: contig of 5157 bp in length
* 37976 38075: gap of 100 bp
* 38076 48059: contig of 9984 bp in length
* 48060 48159: gap of 100 bp
* 48160 58135: contig of 9976 bp in length
* 58136 58235: gap of 100 bp
* 58236 69592: contig of 11357 bp in length
* 69593 69692: gap of 100 bp
* 69693 80834: contig of 11142 bp in length
* 80835 102411: contig of 21477 bp in length
* 102412 102511: gap of 100 bp
* 102512 127924: contig of 25413 bp in length
* 127925 128024: gap of 100 bp
* 128025 162338: contig of 34314 bp in length
* 162339 162438: gap of 100 bp
* 162439 206622: contig of 44184 bp in length.
```

FEATURES

```
source
1..206622
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="11"
  /map="11"
  /clone="RP11-384114"
  /clone_11b="RP11 Human Male BAC"
  1..13492
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left"
misc_feature
13593..14758
  /note="assembly_fragment"
misc_feature
14859..16187
```

```
misc_feature
16288..18628
  /note="assembly_fragment"
misc_feature
18729..22545
  /note="assembly_fragment"
misc_feature
22646..27333
  /note="assembly_fragment"
misc_feature
27434..32718
  /note="assembly_fragment"
misc_feature
32819..37975
  /note="assembly_fragment"
misc_feature
38076..48059
  /note="assembly_fragment"
misc_feature
48160..58135
  /note="assembly_fragment"
misc_feature
58236..69592
  /note="assembly_fragment"
misc_feature
69693..80834
  /note="assembly_fragment"
misc_feature
80935..102411
  /note="assembly_fragment"
misc_feature
102512..127924
  /note="assembly_fragment"
misc_feature
128025..162338
  /note="assembly_fragment"
misc_feature
162439..206622
  /note="assembly_fragment"
  clone_end:T7
  vector_side:right"
BASE COUNT 68410 a 34710 c 35722 g 66279 t 1501 others
ORIGIN
Query Match 35.6%; Score 452.6; DB 2; Length 206622;
Best Local Similarity 80.3%; Pred. No.3 ge-88;
Matches 685; Conservative 0; Mismatches 139; Indels 29; Gaps 12:
QY 269 CCGCGCCCGCGCTTCCAGAGGTTGCTGCTTCCCTGATGCGACGCCAATTAC 328
  || || || || || || || || || || || || || || || || || || || ||
Db 36308 CCCCTTCACACACTCTTAGAGGAGGTACGTGCTACTTCCC-CAGTGTGCGACACTTAC 36366
QY 329 AGCAAAAGCTGGAAAGCTTGCAGCGCACCTTGCAGACACACGGGGAGAACATTGTT 388
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36367 AAAAAGCCCTGGAAGCTTAATGTCACCTGT--AATCACAATGGGGAAGAAACATTGTT 36424
QY 389 TGTGACTATGAGAGGTGTGCGAAGCCCTTCATC-AGGACTACCATCTGAGCGGCAT 447
  || || || || || || || || || || || || || || || || || || || ||
Db 36425 TATGACTGTGAAGGTTGTGCAAGACCCACCTCAAGAGACTAATCATCTGACTATGT 36484
QY 448 TGTGACTCAGACAGAGAAAAGCGTTTGTGTGTCAGCCACTGGCTGTGATC-AAAAAT 506
  || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36485 CTTGA-TCATATGTCAGACAAAGCCATATGTTGTGAGCTAGTAGCTGTGATGTAAAAAT 36543
QY 507 TCAACACAAATTAACACTTGAAGAACATTTTGACGCAACACATGAAATCAACAAAAAC 566
  || || || || || || || || || || || || || || || || || || || ||
Db 36544 TCAACACAAATTAACACTTGAAGAACATTTTGACGCAACACATGAAATCAACAAAAAC 36603
QY 567 A-----ATATATATGCGAGTTTGAAGACTGTGAAGAACCTTTTAAGAAACATGACGCT 621
  || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36604 AAAAATATACATATGCAATTTTGGAGCTTGAAGAACACCTTTTACG-AACATCAGCAGCT 36662
QY 622 GAAATTCATCAGTCCAGCATACCAATGAACCTTATTGAAGTGTACCCAGGAAGATG 681
  || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36663 GAAAAACCATCAGTCCAGCATACCAATGAACCACTTTTCAAGCTGTACCCAGGAAGATG 36722
QY 682 TGGGAACACTTTTGCATCACCAGCAAGCTGAAACGACATGCCAAGGCCACAGAGGCTA 741
  || || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36723 TGGGAACACTTTTCCCATCCATCCAGC-----AAGCGGCAAGGCCATCGGGCTC 36775
QY 742 TGTATGCAAAAAAGATGTTCTTTGGGCAAAAAACATGAGGAACTTGTGAACATGT 801
  || || || || || || || || || || || || || || || || || || || ||
Db 36776 TATATGTCAAAAAGCGTGTCTTTTGTGTGAAGCATGTAGCAAGAACTTTTGAACATGT 36835
QY 802 GAGGAAACCCATTAAGAGGAATACTATGTGAAGTATGCCGGGAAAAACATTTAAACGCA 861
```



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5467230"
/clone_1lb="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
SuperScript II RT (Life Technologies). Note: this is a
NIH_MGC library."
BASE COUNT      301 a      276 c      275 g      223 t      3 others
ORIGIN

```

```

Query Match      72.5%: Score 923; DB 13; Length 1078;
Best Local Similarity 95.2%: Pred. No. 3.8e-192;
Matches 1026; Conservative 0; Mismatches 43; Indels 9; Gaps 7;

```

```

66 GGCCTCGCGCGAAGCTTCAGAGAGGAGCGGCGGCGCGCGGCTCCCGGACAGTG 125
|||||
1 GGCCTCGCGCGAAGCTTCAGAGAGGAGCGGCGGCGCGCGGCTCCCGGACAGTG 60
|||||
126 TCTCGGACGTGGGAGCGCGGCGGCGGCTGGGCTTGGAGGCGCGGCGCGGCTGATCCG 185
|||||
61 TCTCGGACGTGGGAGCGCGGCGGCGGCTGGGCTTGGAGGCGCGGCGCGGCTGATCCG 120
|||||
186 CGGCGGTGCTGCGGAGTCCGCTGCTGCTTGGACATCGCGGACGCTTATTCAGCGG 245
|||||
121 CGGCGGTGCTGCGGAGTCCGCTGCTGCTTGGACATCGCGGACGCTTATTCAGCGG 180
|||||
246 GCGAGAGCTAGCTCCGACCGCGCGCGCGGCTTCCGAGAGGCTTCTGCTGCT 305
|||||
181 GCGAGAGCTAGCTCCGACCGCGCGCGCGGCTTCCGAGAGGCTTCTGCTGCT 240
|||||
306 TCCCTGACTCGACGCGCAATTTACAGCAAGGCTTGAAGCGCGACCTGTCAAGC 365
|||||
241 TCCCTGACTCGACGCGCAATTTACAGCAAGGCTTGAAGCGCGACCTGTCAAGC 300
|||||
366 ACACGGGGGAGAGACATTTGTTGATGAGAGGCTTGGCAAGCGCTTCTCAAGG 425
|||||
301 ACACGGGGGAGAGACATTTGTTGATGAGAGGCTTGGCAAGCGCTTCTCAAGG 360
|||||
426 ACTACATCTGAGCGCGCACATTTCTGACTCAGACAGAGAAAGCGCTTGTGTGAG 485
|||||
361 ACTACATCTGAGCGCGCACATTTCTGACTCAGACAGAGAAAGCGCTTGTGTGAG 420
|||||
486 CCAGTGGCTGTGATCAAAATTTCAACAAATGCAAACTTGAAGAAACATTTTGAACGCA 545
|||||
421 CCAGTGGCTGTGATCAAAATTTCAACAAATGCAAACTTGAAGAAACATTTTGAACGCA 480
|||||
546 AACCTGAAATTCACAAACAAATATATATGACAGTGTGAAGACTGTAAAGACCTTTA 605
|||||
481 AACCTGAAATTCACAAACAAATATATATGACAGTGTGAAGACTGTAAAGACCTTTA 540
|||||
606 AGAATCATGACAGCTGTAATTCATGAGTGCAGCATACCAATGAACCTTATTCAGT 665
|||||
541 AGAATCATGACAGCTGTAATTCATGAGTGCAGCATACCAATGAACCTTATTCAGT 600
|||||
666 GTACCCAGAGAGATGTGGAAACATTTGATCACCAGCAAGCTGAAGAGACATGCCA 725
|||||
601 GTACCCAGAGAGATGTGGAAACATTTGATCACCAGCAAGCTGAAGAGACATGCCA 660
|||||
726 AGGCCCGACAGAGGCTATGATGTGTAAGAGAGTTCCTTTGTGGCAAAACATGACGG 785
|||||
661 AGGCCCGACAGAGGCTATGATGTGTAAGAGAGTTCCTTTGTGGCAAAACATGACGG 720
|||||
786 AACCTTGAATCATGAGAGAAACCATTAAGAGAAATTAATCTATGAGACATGACCGGA 845
|||||
721 AACCTTGTG-AACATGTGAGAGAAACCATTAAGAGAAATTAATCTATGAGACATGACCGGA 779
|||||

```

```

Qy 846 AACATTAAACCCAGAGATTACCTTAAGCAACATGAAAACTCATGCCAGAGAGG 905
|||||
Db 780 NAACATTAAACCCAGAGATTACCTTAAGCAACATGAAAACTCATGCCAGAGAGG 838
|||||
Qy 906 ATGATATCGCTGTCCAGAGAGAGGCTGTGAGACACCTATCTAGTGTATTC 965
|||||
Db 839 ATGATATCGCTGTCCAGAGAGAGGCTGTGAGACACCTATCTAGTGTATTC 898
|||||
Qy 966 AAACCATATCTCTCTTCATGAGGAA-GCGGCGCTTTGTGTGAAACATGCT--G 1022
|||||
Db 899 -AAACCATATCTCTCTTCATGAGGAAAGCGGCGCTTTGTGTGAAACATGCTGG 957
|||||
Qy 1023 GCTGTGCAAAACATTTGCATGAACAAAGTCTCACTAGCATGCTGTGTA--CATGA 1080
|||||
Db 958 CTGTGGCAAAACATTTGCATGAACAAAGTCTCACTAGCATGCTGTGTAACATGAT 1017
|||||
Qy 1081 TCTGTGCAAAAGAAAT-GAAGCTCAAGTCAAAATTCCTGTGAAACAGGAGTTT 1137
|||||
Db 1018 CCTGTGCAAAAGAAATGGAATTCATTCATTCATTCATTCATTCATTCATTCATTC 1075
|||||

```

```

RESULT 2
AL518337      958 bp      mRNA      linear      EST 13-FEB-2001
LOCUS        AL518337 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA009Y018 5
DEFINITION   AL518337 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA009Y018 5
ACCESSION    AL518337
VERSION      AL518337.1 GI:12781830
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 958)
AUTHORS      L.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: seque@genoscope.cns.fr.

```

```

FEATURES
     source
         1..958
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="CS0DA009Y018"
             /clone_1lb="LTI_NFL011_NBC1"
             /sex="male"
             /tissue_type="neuroblastoma cells"
             /lab_host="DH10B"
             /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library is not normalized, but is the control for
the normalized libraries. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
liang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

```

BASE COUNT      270 a      244 c      232 g      195 t      17 others
ORIGIN

```

```

Query Match      72.2%: Score 919; DB 9; Length 958;
Best Local Similarity 97.5%: Pred. No. 2.8e-191;
Matches 935; Conservative 16; Mismatches 6; Indels 2; Gaps 2;

```

```

Qy 76 GAAGTTACAGGAGCGCGGCGCGCGGCTTCCCGGACAGTGTGCGGACG 135
|||||
Db 1 GAAGTTACAGGAGCGCGGCGCGCGGCTTCCCGGACAGTGTGCGGACG 60
|||||
Qy 136 TGGCAGCGCGCTTGGAGGCGCGGCGCGCTTGATCCGCGCGCTGCT 195
|||||

```

```

|||||
Db      61  TGGCAGCGGCGCTGGCGCTTGAGAGCGCGCGCGCTGATCGCGCGCGCTGCT 120
196  CGCGGATGGGTGCTGCTCTTGACCATGCCGCGCTTATGCGACCGCGGAGAGCTC 255
121  CGCGGATGGGTGCTGCTCTTGACCATGCCGCGCGCTTATGCGACCGCGGAGAGCTC 180
256  AGCTCGACCGCGCGCGCGCGCTTGACCATGCCGCGCTTATGCGACCGCGGAGAGCTC 315
181  AGCTCGACCGCGCGCGCGCGCTTGACCATGCCGCGCTTATGCGACCGCGGAGAGCTC 240
316  CAGCGCGCAATTACAGCAAAAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACGCGGGA 375
241  CAGCGCGCAATTACAGCAAAAGCCTGGAAGCTTGAMCGCGCACCTGTGCAAGCACGCGGGA 300
376  GAGACCATTTGTTGTGATGATGAAAGGCTGTGCAAGGCTTATGCGAGCTACATCT 435
301  RANACCATTTGTTGTGATGATGAAAGGCTGTGCAAGGCTTATGCGAGCTACATCT 360
436  GAGCGCGCAATTGCTGATGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 495
361  RAGCGCGCAATTGCTGATGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
496  TGAATCAAAATTCACACAAATCAAACTTGAAGAAACATTTTGAACGCAACATGAAAA 555
421  TGAATCAAAATTCACACAAATCAAACTTGAAGAAACATTTTGAACGCAACATGAAAA 480
556  TCACACAAATTCACATATATATGAGCTTTGAAAGCTGTAAGAACCTTTAAGAACATCA 615
481  TCACACAAATTCACATATATATGAGCTTTGAAAGCTGTAAGAACCTTTAAGAACATCA 540
616  GCAGCTGAAATTCATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 675
541  GCAGCTGAAATTCATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
676  AGATGTTGGGAAACACTTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 735
601  AGATGTTGGGAAACACTTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
736  GGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 795
661  GGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
796  ACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 855
721  ACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
856  ACAGCAAGATTTACCTTAAGCAGACATGAAAGCTGATGCGCCAG - AAAGGATGATGTC 914
781  ACAGCAAGATTTACCTTAAGCAGACATGAAAGCTGATGCGCCAG - AAAGGATGATGTC 840
915  GCGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 974
841  GCGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
975  TCGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
901  TCGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958

```

```

RESULT 3
AL550467 916 bp mRNA linear EST 16-FEB-2001
LOCUS AL550467.L1.NFL006.PL2 Homo sapiens cDNA clone CSDD1057Yn01.5
DEFINITION AL550467.L1.NFL006.PL2 Homo sapiens cDNA clone CSDD1057Yn01.5
ACCESSION AL550467
VERSION AL550467.1 GI:12887469
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 916)

```

```

AUTHORS Li, N.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope - Centre National de Sequencage
Contact: Genoscope
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source 1. 916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD1057Yn01"
/clone.lib="L1.NFL006.PL2"
/tissue.type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@life.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 261 a 238 c 225 g 190 t 2 others
ORIGIN
Query Match 71.0% Score 904: DB 9: Length 916:
Best Local Similarity 99.6% Pred. No. 5, 5e-188:
Matches 904: Conservative 2: Mismatches 2: Indels 0: Gaps 0:
108 GCGGGTCCCGGACGCTGTCGACGTCGACGTCGACGTCGACGTCGACGTCGACGTCG 167
Db 1 GCGGGTCCCGGACGCTGTCGACGTCGACGTCGACGTCGACGTCGACGTCGACGTCG 60
168 CCGGCGCCCTGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 227
Db 61 CCGGCGCCCTGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
228 ACAGCTTATGTCGACGCGCGGAGAGCTGACGTCGACGTCGACGTCGACGTCGACGTC 287
Db 121 ACAGCTTATGTCGACGCGCGGAGAGCTGACGTCGACGTCGACGTCGACGTCGACGTC 180
288 GAGAGTTCATCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 347
Db 181 GAGAGTTCATCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
348 ACAGCTGTCGTCGACGACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407
Db 241 ACAGCTGTCGTCGACGACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
408 GCAAGGCTTCATCAGGAGTACCATCTGAGCGCGCACATTTGATGACTGACAGAGAGAA 467
Db 301 GCAAGGCTTCATCAGGAGTACCATCTGAGCGCGCACATTTGATGACTGACAGAGAGAA 360
468 AGCGGTTGTTGTCGAGCGAGCGAGCGTGTGATCAAAATTCACACAAATTCAAACTGA 527
Db 361 AGCGGTTGTTGTCGAGCGAGCGAGCGTGTGATCAAAATTCACACAAATTCAAACTGA 420
528 AGAATCATTTTGAACGCAATGAAATCAACAAATATATATATGAGTTTGAAG 587
Db 421 AGAATCATTTTGAACGCAATGAAATCAACAAATATATATATGAGTTTGAAG 480
588 ACTGTAAAGAGAGCTTTAAGAAATGAGAGCTGAAATTCATGAGTGCAGAGATGCA 647
Db 481 ACTGTAAAGAGAGCTTTAAGAAATGAGAGCTGAAATTCATGAGTGCAGAGATGCA 540
648 ATGAACCTTATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 707
Db 541 ATGAACCTTATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
708 AGCTGAACGACATGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 767

```

```
Db 601 AGGTGAACACATCCAGAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCCTTTG 660
Qy 768 TGGCAAAAACATGACGAGCACTTCTGAACATGTGAGAGAACCCTAAAGAGAAATAC 827
Db 661 TGGCAAAAACATGACGAGCACTTCTGAACATGTGAGAGAACCCTAAAGAGAAATAC 720
Qy 828 TATGTGAAGTATGCGGAGAAAACATTTAAACGCAAGATACCTTAAACCAACATGAAAA 887
Db 721 TATGTGAAGTATGCGGAGAAAACATTTAAACGCAAGATACCTTAAACCAACATGAAAA 780
Qy 888 CTATATGCCCGCAAGAAAGGATGTATGTCTGTCTCCATGAGAGAAGGCTGTGAGAAACCTATA 947
Db 781 CTATATGCCCGCAAGAAAGGATGTATGTCTGTCTCCATGAGAGAAGGCTGTGAGAAACCTATA 840
Qy 948 CTATGTGTTAAATCTCAAGAGCCATATCTCTCTCTCCATGAGAGAAGGCGCCCTTTTG 1007
Db 841 CAACTGTGTTAAATCTCAAGAGCCATATCTCTCTCTCCATGAGAGAAGGCGCCCTTTTG 900
Qy 1008 TGTGTGAA 1015
Db 901 TGTGTGAA 908

RESULT 4
AL555777 881 bp mRNA linear EST 16-FEB-2001
LOCUS AL555777 LTI_NFL006.Pl2 Homo sapiens cDNA clone CS0DK002YJ18 5
DEFINITION prime, mRNA sequence.
ACCESSION AL555777
VERSION AL555777.1 GI:12897829
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 881)
AUTHORS Li, W.-B., Gruber, C., Jesse, J. and Polyes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL006.Pl2"
/clone="CS0DK002YJ18"
/tissue_type="placenta"
/issue="Vector: PCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime and
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 252 a 229 c 225 g 175 t
ORIGIN
Query Match 69.0%; Score 877.8; DB 9; Length 881;
Best Local Similarity 99.8%; Pired. No. 3e-182;
Matches 879; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 61 CGGCGCTGGCCCTGCGCTTGAGAGCGCGCGCCCTTGATCCGCGCGCTGTGCGCGA 120
Qy 202 GTGGGTGTGCTGCTGTCACATTCGCGCGAGCGGTCATTCGACGCGGAGAGCTCAGCTCC 261
Db 121 GTGGGTGTGCTGCTGTCACATTCGCGCGAGCGGTCATTCGACGCGGAGAGCTCAGCTCC 180
Qy 262 GACCCCGCGCGCGCGCGCTTCCAGAGAGTTTATCTGCTTCCCTGACTGACGCGC 321
Db 181 GACCCCGCGCGCGCGCGCTTCCAGAGAGTTTATCTGCTTCCCTGACTGACGCGC 240
Qy 322 CAATTACAGCAAAAGCTTGAAAGCTTGACGCGCACCTGTGCAAGCACAGGCGGAGAGACC 381
Db 241 CAATTACAGCAAAAGCTTGAAAGCTTGACGCGCACCTGTGCAAGCACAGGCGGAGAGACC 300
Qy 382 ATTTGTTGTGACTATGAAAGGTTGAGAGGCTTCATGAGGACTACATCATGAGACCGG 441
Db 301 ATTTGTTGTGACTATGAAAGGTTGAGAGGCTTCATGAGGACTACATCATGAGACCGG 360
Qy 442 CCACATTCGTGACTCAGACAGAGAGAAAGCGCTTTGTTGTGCAAGCCACTGCTGTGATCA 501
Db 361 CCACATTCGTGACTCAGACAGAGAGAAAGCGCTTTGTTGTGCAAGCCACTGCTGTGATCA 420
Qy 502 AAAATTCACACAAAATCAAACTTTGAAGAAACATTTTGAACGCAAAACATGAAATCAACA 561
Db 421 AAAATTCACACAAAATCAAACTTTGAAGAAACATTTTGAACGCAAAACATGAAATCAACA 480
Qy 562 AAAACATATATATATGCAAGTTTGAAGCTGTGAAGAGACCTTTTAAGAAACATGACAGCT 621
Db 481 AAAACATATATATATGCAAGTTTGAAGCTGTGAAGAGACCTTTTAAGAAACATGACAGCT 540
Qy 622 GAAATTCATCAGTGCAGCATACCAATGAAACCTTATCAAGTGTCCAGAGAGATG 681
Db 541 GAAATTCATCAGTGCAGCATACCAATGAAACCTTATCAAGTGTCCAGAGAGATG 600
Qy 682 TGGGAAACACTTTGTCATCCACGCAAGCTGAAAGCATGCCAGAGCCACAGAGGCTA 741
Db 601 TGGGAAACACTTTGTCATCCACGCAAGCTGAAAGCATGCCAGAGCCACAGAGGCTA 660
Qy 742 TGTATGTCAAAAAGAGTGTCTTGTGGCAAAAACATGTGAGCAACTTGTGAACATGT 801
Db 661 TGTATGTCAAAAAGAGTGTCTTGTGGCAAAAACATGTGAGCAACTTGTGAACATGT 720
Qy 802 GAGGAAACCCATTAAGAGAGAAATACTATGTGAAGTATGCCGCAAAAACATTTAAACGCAA 861
Db 721 GAGGAAACCCATTAAGAGAGAAATACTATGTGAAGTATGCCGCAAAAACATTTAAACGCAA 780
Qy 862 AGATTACTTTAAGACACATGAAAACTCATGCCCCAGAAAGGAGTATGTGCTGCTCC 921
Db 781 AGATTACTTTAAGACACATGAAAACTCATGCCCCAGAAAGGAGTATGTGCTGCTCC 840
Qy 922 AAGGAAAGCTGTGGAAGAACCTATACCTACTGTGTTAATC 962
Db 841 AAGGAAAGCTGTGGAAGAACCTATACCTACTGTGTTAATC 881

RESULT 5
BM562682 1059 bp mRNA linear EST 20-FEB-2002
LOCUS BM562682
DEFINITION AGENCOURT_6588903 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478431
5', mRNA sequence.
ACCESSION BM562682
VERSION BM562682.1 GI:18808978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1059)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
```



```

11:|||||
Db 937 GTRACATATGAGGCTGTGCAAGGCTTCATCAGGAGTACATCTGAGCCGACATTC 878
Qy 450 TGCATCAGCAGAGAGAAAGCGTTGTTGTGCAAGCCACTGGCTGTATCAAAATTA 509
Db 877 TTAATCAGCAGAGAGAAAGCGTTGTTGTGTC -GCCAATGGCTGTATCAAAATTA 819
Qy 510 ACACAAATCAAACTTGAAGAAATTTGAAGCGCAACATGAAATCAACAAACAAT 569
Db 818 ACACAAATCAAACTTGAAGAAATTTGAAGCGCAACATGAAATCAACAAACAAT 759
Qy 570 ATATATGCACTTTGAAGAGCTGTAGAGAGCTTTAAGAAATCAGCAGCTGAATCC 629
Db 758 ATATATGCACTTTGAAGAGCTGTAGAAAGACCTTTAAGAAATCAGCAGCTGAATCC 699
Qy 630 ATCAGTGCAGATACCATTAAGCTCTATTCAGTGTACCAAGGAAAGATGGGAAAC 689
Db 698 ATCAGTGCAGATACCATTAAGCTCTATTCAGTGTACCAAGGAAAGATGGGAAAC 639
Qy 690 ACTTTCATCACCAGCAGAGCTGAAAGGACATGCCAAGCCGACGAGGCTATGTATGTC 749
Db 638 ACTTTCATCACCAGCAGAGCTGAAAGGACATGCCAAGCCGACGAGGCTATGTATGTC 579
Qy 750 AAAAGAGATGTCCTTTGTGGCAAAACATGAGAGACTTGTGAACATGTGAGAGAA 809
Db 578 AAAAGAGATGTCCTTTGTGGCAAAACATGAGAGACTTGTGAACATGTGAGAGAA 519
Qy 810 CCCATTAAGAGAAATGATGTAAGTGTGAGTGTGCGGAAACATTTAAACGAAAGATTAC 869
Db 518 CCCATTAAGAGAAATGATGTAAGTGTGAGTGTGCGGAAACATTTAAACGAAAGATTAC 459
Qy 870 TTAAAGCAACATGAAACATCATGCCCCAGAAAGAGATGTATGTCCTGCAAGAGAG 929
Db 458 TTAAAGCAACATGAAACATCATGCCCCAGAAAGAGATGTATGTCCTGCAAGAGAG 399
Qy 930 GCTGTGGAAGAACTTACTACTGTGTTAAATCTCCAAAGCATATCTCTCTCCATG 989
Db 398 GCTGTGGAAGAACTTACTACTGTGTTAAATCTCCAAAGCATATCTCTCTCCATG 339
Qy 990 AGGAAGCGGCGCTTTGTGTGTGAACATGCGGCTGTGGCAAAACATTTGCAATGAAC 1049
Db 338 AGGAAGCGGCGCTTTGTGTGTGAACATGCGGCTGTGGCAAAACATTTGCAATGAAC 279
Qy 1050 AAAGTCTCACTAGGAGTGTGTTGATCATGATCTGTACAAAGAAATGAAGCTTCAAG 1109
Db 278 AAAGTCTCACTAGGAGTGTGTTGATCATGATCTGTACAAAGAAATGAAGCTTCAAG 219
Qy 1110 TCAGAAATCTGTGTAAGAAAGGAGTTGGCTCTCATCTCAAGTGAATATCTCCCA 1169
Db 218 TCAGAAATCTGTGTAAGAAAGGAGTTGGCTCTCATCTCAAGTGAATATCTCCCA 159
Qy 1170 AAAGGAAGAGGCGGAGGCTTATCTTTGTGCAAAAGGAGAGTACCCCACTGTGG 1229
Db 158 AAAGGAAGAGGCGGAGGCTTATCTTTGTGCAAAAGGAGAGTACCCCACTGTGG 99
Qy 1230 AAGACAGATGCTCTGAGAGTGTGAGTACTTACCTTGGCTAA 1273
Db 98 AAGACAGATGCTCTGAGAGTGTGAGTACTTACCTTGGCTAA 55

```

```

RESULT 7
LOCUS B0650498 930 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8207703 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283123
5', mRNA sequence.
ACCESSION B0650498
VERSION B0650498.1 GI:21774670
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)

```

```

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LICM2478 row: 1 column: 20
High quality sequence stop: 636.
FEATURES
source
1. 930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6283123"
/clone_lib="NIH-MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/note="Organ: Liver; Vector: pOTB1; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAC(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."
BASE COUNT 316 a 212 c 200 g 201 t 1 others
ORIGIN
Query Match 67.6%; Score 860.4; DB 14; Length 930;
Best Local Similarity 97.0%; Pred. No. 2e-178;
Matches 899; Conservative 0; Mismatches 22; Indels 6; Gaps 2;
Qy 301 CTCCTCCCTGAGTCGAGGCGCAATTAAGCAAGGCTGGAAGCTTGAGCGGACCTGTG 360
Db 1 CTCCTCCCTGAGTCGAGGCGCAATTAAGCAAGGCTGGAAGCTTGAGCGGACCTGTG 60
Qy 361 CAAGCAGCGGGGAGAGACCATTTCTTTGTGACTATGAGGCTGTGGCAAGGCTTCAT 420
Db 61 CAAGCAGCGGGGAGAGACCATTTCTTTGTGACTATGAGGCTGTGGCAAGGCTTCAT 120
Qy 421 CAGGACTACCATCTGAGCGCCACATTTCTGACTCAGACAGAGAAAGCCGTTGTTG 480
Db 121 CAGGACTACCATCTGAGCGCCACATTTCTGACTCAGACAGAGAAAGCCGTTGTTG 180
Qy 481 TGCAGCAGTGGCTGTGATCAAAATTCACCAAAATTAACCTGAAGAAATTTGA 540
Db 181 TGCAGCAGTGGCTGTGATCAAAATTCACCAAAATTAACCTGAAGAAATTTGA 240
Qy 541 ACGCAAAACATGAAATCAACAAAAACATATATATGACATTTTGAAGAGTGAAGAG 600
Db 241 ACGCAAAACATGAAATCAACAAAAACATATATATGACATTTTGAAGAGTGAAGAG 300
Qy 601 CTTTAAAGAAACATCAGCAGCTGAAATCCATGACGATACCATGAATCACTTAT 660
Db 301 CTTTAAAGAAACATCAGCAGCTGAAATCCATGACGATACCATGAATCACTTAT 360
Qy 661 CAAAGTACCCAGGAAGATGGGAAACACTTTGATGATCACCAGCAACCTGAAGAG 720
Db 361 CAAAGTACCCAGGAAGATGGGAAACACTTTGATGATCACCAGCAACCTGAAGAG 420
Qy 721 TCCCAAGGCCCAAGAGGCTATGTATGTCAAAAAGGATTTCTTTGGCAAAAACATG 780
Db 421 TCCCAAGGCCCAAGAGGCTATGTATGTCAAAAAGGATTTCTTTGGCAAAAACATG 480
Qy 781 GACGGAATCTTGAACATGTAGAGACAAACCATTAAGAGAAATCTATGTGAAGTATG 840
Db 481 GACGGAATCTTGAACATGTAGAGACAAACCATTAAGAGAAATCTATGTGAAGTATG 540

```

QY	841	CCGGAACATTTTAAACCAAGAATTACCTTACGACATGAAGAAACTCATGCCCCAGA	900
Db	541	CCGGAACATTTTAAACCAAGAATTACCTTACGACATGAAGAAACTCATGCCCCAGA	600
QY	901	AAGGATGTATGTGCGTGTCCAGAAGAGCGTGTGGAGAAACCTATACTGTTTAA	960
Db	601	AAGGATGTATGTGCGTGTCCAGAAGAGCGTGTGGAGAAACCTATACTGTTTAA	660
QY	961	TCCTCAAGGCATATCCTCTCTCTTCATGAGAGAAAGCCGCTTTTGTGTGAACATGC	1020
Db	661	TCCTCAAGGCATATCCTCTCTCTTCATGAGAGAAAGCCGCTTTTGTGTGAACATGC	720
QY	1021	TGGCTGTGGCAAAACATTTGCAATGAATGAAGTCACCTACTAGCGTGGTGTGCATGCA	1080
Db	721	TGGCTGTGGCAAAACATTTGCAATGAATGAAGTCACCTACTAGCGTGGTGTGCATGCA	780
QY	1081	TCCTGACAGAAGAAATGAAGCTCAAGTCAAAAATCTCGTGAAGAAACGAGTTTGGC	1140
Db	781	TCCTGACAGAAGAAATGAAGCTCAAGTCAAAAATCTCGTGAAGAAACGAGTTTGGC	840
QY	1141	CTCCATCTCTAGTGTG--ATATATCCCTCCCAAGGAAACAAAGGCAAGG--CTATTC	1194
Db	841	CTCCATCTCTAGTGTG--ATATATCCCTCCCAAGGAAACAAAGGCAAGGCTCATCT	900
QY	1195	TTTGTGTCAAAACGAGAGTCAACCAA	1221
Db	901	TTTGTGTCAAAACGAGAGTCAACCCA	927
RESULT 8			
LOCUS	AL527493	957 bp	mRNA
DEFINITION	AL527493	lTL1_NFL003_NBC3	Homo sapiens cDNA clone CS00C024YE05
VERSION	AL527493		prime, mRNA sequence.
KEYWORDS	AL527493.1	GI:12790986	EST.
SOURCE			human.
ORGANISM			Homo sapiens
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS			Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.
TITLE			Full-length cDNA libraries and normalization
JOURNAL			Unpublished (2001)
COMMENT			Contact: Genoscope
			Genoscope - Centre National de Sequencage
			BP 191 91006 Evry cedex - France
			Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES			Location/Qualifiers
source			1..957
			/organism="Homo sapiens"
			/db_xref="taxon:9606"
			/clone="CS00C024YE05"
			/clone_id="lTL1_NFL003_NBC3"
			/sex="male"
			/tissue_type="neuroblastoma cells"
			/lab_host="DH10B"
			/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng liang life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8372 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"
BASE COUNT	283 a	240 c	227 g 206 t 1 others
ORIGIN			
Query Match		66.7%	Score 849.4; DB 9; Length 957;

	Best Local Similarity 91.9%; Pred. No. 5,26-176;	
	Matches 948; Conservative 1; Mismatches 7; Indels 75; Gaps 2	
OY	115 CCGGCAAGTGTCTCGGCACGTGGACGCGCGCTTGCCCTGGGCTTTGSAGCGCGGCGC 174	
Db	1 CCGGCAAGTGTCTCGGCACGTGGACGCGCGCGCTTGCCCTGGGCTTTGSAGCGCGGCGC 60	
OY	175 CCTGGATCCGCCGCGCGCTGTCGCCGAGTGGGTGTCTGTCTTGACACATCGCCGACGGTT 234	
Db	61 CCTGGATCCGCCGCGCGCTGTCGCCGAGTGGGTGTCTGTCTTGACACATCGCCGACGGTT 120	
OY	235 CATTT-GAAGCGGCGAGACTCAGCTCCGACCAGCCCCGCGCGCGCGCTTCCAGAGGT 293	
Db	121 CATTTAGCAGCGCGCGAGACTCAGCTCCBACCCCGCGCGCGCTTCCAGAGAGT 180	
OY	294 TCATTGCTCTTCCCTGACTCGACGCGCATTTACAGCAAAGCCTGGAAAGCTTGACGCGC 353	
Db	181 TCAATGCTCTCTTCCCTGACTCGAGCGCATTTACAGCAAAGCCTGGAAAGCTTGACGCGC 240	
OY	354 ACCGTGTCGAAGCACACGCGGGAGAGAACATTTGTTGTGACTATGAAAGGTGTGGCAAG 413	
Db	241 ACCTGTGCAAGCACACGCGGGAGAGACCAATTTGTTGTGACTATGAAAGGTGTGGCAAG 300	
OY	414 CCTTCATCAGGGGCTACCATCTGAGCGCGCAATTTCTGACTCACACAGAGAAAGCGGT 473	
Db	301 CCTTCATCAGGGGCTACCATCTGAGCGCGCAATTTCTGACTCACACAGAGAAAGCGGT 360	
OY	474 TTGTTGTGCAAGCTGCGCTGTGATCAAAAATTCACACAATAATCAAACTTGANGAAC 533	
Db	361 TTGTTGTGCAAGCTGCGCTGTGATCAAAAATTCACACAATAATCAAACTTGANAAGAC 420	
OY	534 ATTTTGAAGCAAAACATGAAAAATCAACAAAACAAATATATATGCAAGTTTTGAGAAGCTGA 593	
Db	421 ATTTTGAAGCAAAACATGAAAAATCAACAAAACAAATATATATGCAAGTTTTGAAAGCTGA 480	
OY	594 AGAAGACCTTTAAGAAACATCAGACGCTGAAAAATCATCAGTCGACAGATACCAATGAAC 653	
Db	481 AGAAGACCTTTAAGAAACATCAGACGCTGAAAAATCATCAGTCGACAGATACCAATGAAC 540	
OY	654 CTCTATTCAAGTGTACCCAGAGAAAGATGTGGAAACACTTTGATCAACCCAGCAAGCTGA 713	
Db	541 CTCTATTCAAG-----GCTATGTATGTCAAAAAGAGATGTTCTTGTGGCAA 550	
OY	714 AAGGACATGGCCAAGGCCACGAGGCGTATGTATGTCAAAAAGAGATGTTCTTGTGGCAA 773	
Db	551 -----GCTATGTATGTCAAAAAGAGATGTTCTTGTGGCAA 586	
OY	774 AAACATGAGAGGAACCTTGTGAAMACNTGTGAGAGAAACCCATTAAGAGAGAAATACTATGTG 833	
Db	587 AAACATGAGAGGAACCTTGTGAAMACNTGTGAGAGAAACCCATTAAGAGAGAAATACTATGTG 646	
OY	834 AAGATATCCGGAAMAACATTTTAAACGCAAGATTTACCTTAAGCAACACATGAAAACTATG 893	
Db	647 AAGATATCCGGAAMAACATTTTAAACGCAAGATTTACCTTAAGCAACACATGAAAACTATG 706	
OY	894 CCCGAGAAAGGATGTATGTCTGCTGTCCAAGAGAAAGGCTGTGGAGAAACCTATACTACTG 953	
Db	707 CCCGAGAAAGGATGTATGTCTGCTGTCCAAGAGAAAGGCTGTGGAGAAACCTATACTACTG 766	
OY	954 TGTTTAATTCGCAAAAGCCATATCTCTCTTCATGAGAGAAAGCGCCCTTTTGTGTGTG 1013	
Db	767 TGTTTAATTCGCAAAAGCCATATCTCTCTTCATGAGAGAAAGCGCCCTTTTGTGTGTG 826	
OY	1014 AACATGCTGTGCTGTGGCAAAAACATTTTGCACATGAAACAAAGTCTCACTAGGAGATCCTGTTG 1073	
Db	827 AACATGCTGTGCTGTGGCAAAAACATTTTGCACATGAAACAAAGTCTCACTAGGAGATCCTGTTG 886	
OY	1074 TACATGATCTGTGACAAGAAAGAAATGAAAGCTCAAAAGTCAAAAATATCTGTGAAAAACGGA 1133	
Db	887 TACATGATCTGTGACAAGAAAGAAATGAAAGCTCAAAAGTCAAAAATATCTGTGAAAAACGGRG 946	
OY	1134 GTTTGGGCTCT 1144	

Db	947	TTTGGCCCT	957	
RESULT 9				
LOCUS	AL574698		906 bp	mrna
DEFINITION	AL574698 LTL.NFL006.PL2 Homo sapiens cDNA clone CS001065Y123			linear EST 16-FEB-2001
ACCESSION	AL574698			
VERSION	AL574698.1	GI:12935148		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 906)			
TITLE	L.M.B., Gruber,C., Jesssee,J. and Polayes,D.			
JOURNAL	Full-length cDNA libraries and normalization			
COMMENT	Unpublished (2001)			
FEATURES	Feature: Genoscope			
	Genoscope - Centre National de Sequencage			
	BP 191 91006 EVRY cedex - France			
	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.			
	Location/Qualifiers			
	1. 906			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="CS001065Y123"			
	/clone_lib="LTL.NFL006.PL2"			
	/issue_type="placenta"			
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"			
BASE COUNT	196 a	173 c	199 g	309 t
ORIGIN			29 others	
Query Match	66.3%	Score 844.6	DB 9	Length 906
Best Local Similarity	95.1%	Prod. No. 5.8e-175		
Matches 840	Conservative 17	Mismatches 17	Indels	Gaps 0
QY	391 TGACATGATGAAGGCTTGCGCAAGCCCTTCATCATCGGGAGCTACATCTGAGCGCCACATTCT	450		
DB	906 TTTACATATGATGAAGGCTTGCGCAAGCCCTTCATCATCGGGAGCTACATCTGAGCGCCACATTCT	847		
QY	451 GACTCAGCAGGAGAAAAGCCGTTTGTGTCGACAGCCACATGCGTGTGATCAAAAATTCAA	510		
DB	846 GACTCAGCAGGAGAAAAGCCGTTTGTGTCGACAGCCACATGCGTGTGATCAAAAATTCAA	787		
QY	511 CACAAATATCAAACTTGAGAAACATTTTGAACGCGAATCAATGAATCAACAAAACATA	570		
DB	786 CACAAATATCAAACTTGAGAAACATTTTGAASGCAACATATGAATAATSAACAAAASATA	727		
QY	571 TATATGACGTTTGAAGACTGTACAAAGCCCTTAAAGAAACATCGACGAGCCGTAATAATCA	630		
DB	726 TATATGACGTTTGAAGACTGTACAAAGCCCTTAAAGAAACATCGACGAGCCGTAATAATCA	667		
QY	631 TCAGTGCCAGCATATCAATGAACCTCTATATTCAGTGTACCCAGGAAGATGTGGGAAACA	690		
DB	666 TCAGTGCCAGCATATCAATGAACCTCTATATTCAGTGTACCCAGGAAGATGTGGGAAACA	607		
QY	691 CTTTGATATCACCAGCAAGCTGAAGACATGCCAAGGCCACGAGGCCATATGATGTCA	750		
DB	606 CTTTGATATCACCAGCAAGCTGAAGACATGCCAAGGCCACGAGGCCATATGATGTCA	547		
QY	751 AAAAGATGTTCTTTGGCAAAAACATGAGCGAATCTTGAAACATGTAGAGAAC	810		

D	b	546	AAGAGATGTTCTTGTGGCGAATAACATTCGACGAACACTTCTGAAACATGTGAGGAAC	487
O	y	811	CCATAAGAGGAATATCATGTGAAGTATGCCGGAAAAATTAAACCAAGATTACT	870
D	b	486	CCATTAAGAGGAATATCATGTGAAGTATGCCGGAAAAATTAAACCAAGATTACT	427
O	y	871	TAAACAACATGAAAATCTCATGCCCGCAAAAGGATATGTGCCTGCCAACGAAG	930
D	b	426	TAAAGAACAATGAAAATCTCATGCCCGCAAAAGGATATGTGCCTGCCAACGAAG	367
O	y	931	CTGTGAGAAACCTTACTACTGTGTAAATCTCCAAAGCATTATCCTCTCCATGA	990
D	b	366	CTSTBAAGAACCTTAACAACCTSTGTTTAATCTCCMAAACATATCCCTTMCATGA	307
O	y	991	GGAAGACCCCTTTTGTTGTGTGAACATGCTGGCTGTGGCAAAACATTTGCAATGA	1050
D	b	306	GGAAGACCCCTTTTGTTGTGTGAACATGCTGGCTGTGGCAAAACATTTGCAATGA	247
O	y	1051	AAGTCTCACTAGGCATGCTGTTGTATCATGATCTGACAAAGAAATGAAGCTCAA	1110
D	b	246	AAATCTCACTAGGCATGCTGTTGTATCATGATCTGACAAAGAAATGAAGCTCAA	187
O	y	1111	CAAAAATCTGCTGTAATAACGAGATTTGGCCCTCATCTCAGTGGATATATCCGCC	1170
D	b	186	YAAAAATCTGCTTKAAAAACGAGATTTGGCCCTCATCTCAGTGGATATATCCGCC	127
O	y	1171	AAGGAACAAGGCGCAAGGCTTATCTTTGTGTCAAAACGAGAGTACCACATGTGTGA	1230
D	b	126	AAGGAACAAGGCGCAAGGCTTATCTTTGTGTCAAAACGAGAGTACCACATGTGTGA	67
O	y	1231	AGACAAGATGCTCTTCGACAGTTGCAGTACTTAACCTTGCGCTAA	1273
D	b	66	AGACAAGATGCTCTTCGACAGTTGCAGTACTTAACCTTGCGCTAA	24

RESULT 10
BM541851

LOCUS BM541851 1111 bp mRNA linear EST 20-FEB-2002

DEFINITION AGNCOURT_6439143 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5531503 5', mRNA sequence.

ACCESSION BM541851

VERSION BM541851.1 GI:18770849

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS Contact: Robert Strausberg, Ph.D.
TITLE Email: cgabbs-ref@mail.nih.gov
JOURNAL Tissue Procurement: ATCC

COMMENT CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LHAM1213 row: e column: 08
High quality sequence stop: 654.

FEATURES
source location/Qualifiers
1..1111 /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5531503"
/clone_id="NIH_MGC_71"
/tissue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."


```

|||||
Db 361 TCAGGAGCTACATCTGAGCGCCGACATCTGACTCAGACAGAGAAAGCCGTTGTTT 420
QY 480 GTGACGACCTGCTGCTGATCAAAAATTCACACACAAAATTAACCTGACAAACATTTTG 539
Db 421 GTGACGACCAATGGCTGTGATCAAAAATTCACACACAAAATTAACCTGACAAACATTTTG 480
QY 540 AACGCAAAATGAAAAATCAACAAAACAAATATATATGTCAGTTTGAAGACTGTAGAGA 599
Db 481 AACGCAAAATGAAAAATCAACAAAACAAATATATATGTCAGTTTGAAGACTGTAGAGA 540
QY 600 CCTTTAAGAAATCATCAGACCTGAAAAATTCATGAGTCCAGCATATCAATGAACTCTAT 659
Db 541 CCTTTAAGAAATCATCAGACCTGAAAAATTCATGAGTCCAGCATATCAATGAACTCTAT 600
QY 660 TCAGCTTACCCAGAGAGATGTGGGAAACACTTTGATACCCAGCAACTGAAAGAC 719
Db 601 TCAGCTTACCCAGAGAGATGTGGGAAACACTTTGATACCCAGCAACTGAAAGAC 660
QY 720 ATCCCAAGCCCAAGAGGCTATGATGTCAAAAAGAGATTTCTTTGTGGCAAAAACAT 779
Db 661 ATCCCAAGCCCAAGAGGCTATGATGTCAAAAAGAGATTTCTTTGTGGCAAAAACAT 720
QY 780 GAGCGAATCTTGAAACATGTGAGAGAAACCATAAAGAGAAATCTATGTGAAGTAT 839
Db 721 GAGCGAATCTTGAAACATGTGAGAGAAACCATAAAGAGAAATCTATGTGAAGTAT 780
QY 840 CCGCGAAACATTTAAAGCGAAAGATTACCTTAAGCAACATGAAACATCATGCCAG 899
Db 781 CCGCGAAACATTTAAAGCGAAAGATTACCTTAAGCAACATGAAACATCATGCCAG 840
QY 900 AAAGCG 905
Db 841 AAANCG 846

```

```

RESULT 12
BO644275 902 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_835851 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285421
DEFINITION 5', mRNA sequence.
ACCESSION BO644275
VERSION BO644275.1 GI:21768447
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgepds-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1CM2484 row: j column: 14
High quality sequence stop: 642.
Location/Qualifiers
1..902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6285421"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

```

FEATURES
source
1..902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6285421"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

```

BASE COUNT 301 a 208 c 197 g 194 t 2 others
ORIGIN
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1 kbp. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."
Query Match 65.7%; Score 836; DB 14; Length 902;
Best Local Similarity 98.3%; Pred. No. 4,5e-173;
Matches 875; Conservative 0; Mismatches 12; Indels 3; Gaps 3:
QY 300 GCTCCCTCCGCTGAGGCGCAATTCAGCAAGCGCTGGAAGCTTGAGCGGCACTGT 359
Db 11 GCTCCCTCCGCTGAGGCGCAATTCAGCAAGCGCTGGAAGCTTGAGCGGCACTGT 70
QY 360 GCAAGCAGACGGGGGAGAGACATTTGTTGTGACTATGAAGGGTGTGCAAGGCTTCA 419
Db 71 GC-ANCAAGCGGGGAGAGACATTTGTTGTGACTATGAAGGGTGTGCAAGGCTTCA 129
QY 420 TCAGGACTACATCTGAGCGCCACATCTGACTCACAGAGAAACCCCTTTGTTT 479
Db 130 TCAGGAGCTACATCTGAGCGCCACATCTGACTCACAGAGAAACCCCTTTGTTT 189
QY 480 GTGAGCGCACTGGCTGTGATCAAAAATTCACACAAAATCAAACTTGAACATTTTG 539
Db 190 GTGAGCGCAATGGCTGTGATCAAAAATTCACACAAAATCAAACTTGAACATTTTG 249
QY 540 AACGCAAAATGAAAAATCAACAAAACAAATATATGTCAGTTTGAAGACTGTAGAGA 599
Db 250 AACGCAAAATGAAAAATCAACAAAACAAATATATGTCAGTTTGAAGACTGTAGAGA 309
QY 600 CCTTTAAGAAATCATCAGACCTGAAAAATCCATCAGTCCAGCATATCAATGAACTCTAT 659
Db 310 CCTTTAAGAAATCATCAGACCTGAAAAATCCATCAGTCCAGCATATCAATGAACTCTAT 369
QY 660 TCAGCTTACCCAGAGAGATGTGGGAAACACTTTGATACCCAGCAACTGAAAGAC 719
Db 370 TCAGCTTACCCAGAGAGATGTGGGAAACACTTTGATACCCAGCAACTGAAAGAC 429
QY 720 ATCCCAAGCCCAAGAGGCTATGATGTCAAAAAGAGATTTCTTTGTGGCAAAAACAT 779
Db 430 ATCCCAAGCCCAAGAGGCTATGATGTCAAAAAGAGATTTCTTTGTGGCAAAAACAT 489
QY 780 GAGCGAATCTTGAAACATGTGAGAGAAACCATAAAGAGAAATCTATGTGAAGTAT 839
Db 490 GAGCGAATCTTGAAACATGTGAGAGAAACCATAAAGAGAAATCTATGTGAAGTAT 549
QY 840 GCGGAAACATTTAAAGCGAAAGATTACCTTAAGCAACATGAAACATCATGCCAG 899
Db 550 GCGGAAACATTTAAAGCGAAAGATTACCTTAAGCAACATGAAACATCATGCCAG 609
QY 900 AAAGGATGTATGCTGCTGTCACAAAGAGAGCTGTGGAAGACCTATCTACTGTGTTA 959
Db 610 AAAGGATGTATGCTGCTGTCACAAAGAGAGCTGTGGAAGACCTATCTACTGTGTTA 669
QY 960 ATCTCCAAAGCATATCTCTCTCTCCATGAGAAACCCCTTTTGTGTGGAACATG 1019
Db 670 ATCTCCAAAGCATATCTCTCTCTCCATGAGGAAACCCCTTTTGTGTGGAACATG 729
QY 1020 CTGGCTGTGGCAAAACATTTGCAATGAAGAAACAGTCTCATGAGCAATGCTGTGATC 1079
Db 730 CTGGCTGTGGCAAAACATTTGCAATGAAGAAACAGTCTCATGAGCAATGCTGTGATC 789
QY 1080 ATCTGCAAGAAAGAAATGAAAGCTCAAAAATATCTGTAAGAAAGAGGATTTG 1139
Db 790 ATCTGCAAGAAAG -AAATGAAGCTCAATCCAAAATATCTGTAAGAAAGAGGATTTG 848
QY 1140 CCTCTCATCTCAGTGAATATAT -CCCTCCGCAAAAGAGAAAGAGGCAAG 1188
Db 849 CCTCTCATCTCAGTGAATATATCCCTCCGCAAAAGAGAAAGAGGCAAG 898

```

```

RESULT 13
AL578193/c
LOCUS      AL578193      933 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION AL578193 L1L_NFL006_P12 Homo sapiens cDNA clone CS0DK002Y018 3
prime, mRNA sequence.
ACCESSION  AL578193
VERSION     AL578193.1  GI:12942034
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 933)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CS0DK002Y018"
/tissue_type="placenta"
/Note="Vector: PCWVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      202 a      192 c      207 g      326 t      6 others

Query Match      65.2%; Score 830.2; DB 9; Length 933;
Best Local Similarity 98.5%; Pred. No. 8.3e-172;
Matches 854; Conservative 4; Mismatches 7; Indels 2; Gaps 2;

QY 408 GCAAGCCCTTCATCAGGACTACCATGTGAGCGCCACATTTGACACACAGAGAAA 467
|||||
Db 933 GCAAGGCCCTTCATCA-GGACTACCATCTGAGCGCCACATTTGACACACAGAGAAA 875

QY 468 AGCGGTTGTTGTGCAAGCCAGCTGGCTGATCAAAATTCACACAAATCAAACTTGA 527
|||||
Db 874 AGCGGTTGTTGTGCAAGCCAGTGGCTGTGATCAAAATTCACACAAATCAAACTTGA 815

QY 528 AGAATCATTTTGAAGCAGCAATGAAATCAACAAATATATATATGAGTTTGAAG 587
|||||
Db 814 AGAATCATTTTGAAGCAGCAATGAAATCAACAAATATATATATGAGTTTGAAG 755

QY 588 ACTGTAAAGAGACCTTTAAGAACATGACAGCTGAAATTCATCAGTCCAGATACCA 647
|||||
Db 754 ACTGTAAAGAGACCTTTAAGAACATGACAGCTGAAATTCATCAGTCCAGATACCA 695

QY 648 ATGAACCTGATTCAGTGAACCGAGAGAGTGTGGAACATTTGCAATCACCAGCA 707
|||||
Db 694 ATGAACCTGATTCAGTGAACCGAGAGAGTGTGGAACATTTGCAATCACCAGCA 635

QY 708 AGCTGAAGCAGCATGCCAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCTTTG 767
|||||
Db 634 AGCTGAAGCAGCATGCCAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCTTTG 575

QY 768 TGGCAAAAACATGAGCGAATCTTGAACATGTGAGAGAAACCCATPAAGAGGAAATAC 827
|||||
Db 574 TGGCAAAAACATGAGCGAATCTTGAACATGTGAGAGAAACCCATPAAGAGGAAATAC 515

```

```

QY 828 TATGTGAAGTATGCCGGAACAAATTTAAACGCAAGATTACTTTAAGCAACATGAAA 887
|||||
Db 514 TATGTGAAGTATGCCGGAACAAATTTAAGCAAGATTACTTTAAGCAACATGAAA 435

QY 888 CTCATGCCCGCAGAAAGGAGTGTATGTCGTCCAGAGAGAGGCTGTGAAGAACCATA 947
|||||
Db 454 CTCATGCCCGCAGAAAGGAGTGTATGTCGTCCAGAGAGAGGCTGTGAAGAACCATA 395

QY 948 CTACTGTGTTTAATCTCCAAAGCCATATCCTCTCCTTCATGAGAGAAAGCCCTTTTG 1007
|||||
Db 394 CAACCTGTGTTTAATCTCCAAAGCCATATCCTCTCCTTCATGAGAGAAAGCCCTTTTG 335

QY 1008 TGTGTGAACATGCTGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTAGCATG 1067
|||||
Db 334 TGTGTGAACATGCTGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTAGCATG 275

QY 1068 CTGTTGTACATGATCCTGACAGAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGA 1127
|||||
Db 274 CTGTTGTACATGATCCTGACAGAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGA 215

QY 1128 AACGGAATTTGGCTCTCATCTCAGTGTATATCCTCCCAAAAGAAACAAAGGCAAG 1187
|||||
Db 214 AACGGAATTTGGCTCTCATCTCAGTGTATATCCTCCCAAAAGAAACAAAGGCAAG 155

QY 1188 GCTTATCTTTGTGCAAAAGGAGAGTCAACCAACTGTGTGAAGACATAGTCTCGA 1247
|||||
Db 154 GCTTATCTTTGTGCAAAAGGAGAGTCAACCAACTGTGTGAAGACATAGTCTCGA 95

QY 1248 CAGTTGCA-GTACTTACCCTTGCTAA 1273
|||||
Db 94 CAGTTGCAAGCACTTACCCTTGCTAA 68

```

```

RESULT 14
BQ278693      1089 bp      mRNA      linear      EST 07-MAY-2002
LOCUS      BQ278693
DEFINITION AGENCOURT 7049694 NIH_MGC_107 Homo sapiens cDNA clone IMAGE:5805333
5', mRNA sequence.
ACCESSION  BQ278693
VERSION     BQ278693.1  GI:20488901
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1089)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
Email: cgapubs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1002044 row: f column: 22
High quality sequence stop: 610.
Location/Qualifiers
1. 1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5805333"
/clone_lib="NIH_MGC_107"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(g). Library constructed by

```

Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library.

BASE COUNT 296 a 289 c 283 g 221 t

Query Match 65.2%; Score 830; DB 14; Length 1089;

Best Local Similarity 94.6%; Pred. No. 9.3e-172;

Matches 927; Conservative 0; Mismatches 40; Indels 13; Gaps 6;

```

OY 70 TCCGCCAGAGTTTCAGCAGGAGCCGTGGCCGCGCGCGCTTCCGCGACGCTGCTC 129
    |||||
Db 1 TCCGCGCAAGTTTCAGCAGGAGCCGTGGCCGCGCGCGCTTCCGCGACGCTGCTC 60
    |||||

OY 130 GGCACGTGGAGGGCGCGCTGGCCCTGGAGGCGCGCGCGCTTCCGATCCCGCGC 189
    |||||
Db 61 GGCACGTGGAGGGCGCGCTGGCCCTGGAGGCGCGCGCGCTTCCGATCCCGCGC 120
    |||||

OY 190 CGTGTGCGCGAGTGGTGTCTCTCTTGCACCATGCGCGAGCGCTTCCAGCGCGCA 249
    |||||
Db 121 CGTGTGCGCGAGTGGTGTCTCTTGCACCATGCGCGAGCGCTTCCAGCGCGCA 180
    |||||

OY 250 GAGCTGAGCTCCGACCCCGCGCGCGCGCGCTTCCAGAGGTTCACTGCTCTTCCC 309
    |||||
Db 181 GAGCTGAGCTCCGACCCCGCGCGCGCGCGCTTCCAGAGGTTCACTGCTCTTCCC 240
    |||||

OY 310 TGACTGAGCGCCAAATTACAGCAAAAGCTGGAAGCTTGACGCGCACTTGCAAGCAAC 369
    |||||
Db 241 TGACTGAGCGCCAAATTACAGCAAAAGCTGGAAGCTTGACGCGCACTTGCAAGCAAC 300
    |||||

OY 370 GGGGAGAGACCAATTTGTGTGACTATGAAGGCTGTGGCAAGCGCTTCAATCAGGACTA 429
    |||||
Db 301 GGGGAGAGACCAATTTGTGTGACTATGAAGGCTGTGGCAAGCGCTTCAATCAGGACTA 360
    |||||

OY 430 CCATCTAGCGCGCACATTTCTGACTCAGACAGAGAAAAGCCGTTTGTGTGACGCCAC 489
    |||||
Db 361 CCATCTAGCGCGCACATTTCTGACTCAGACAGAGAAAAGCCGTTTGTGTGACGCCAA 420
    |||||

OY 490 TGGCTGTGATCAAAAATTTCAACAACAAATCAAACTTGAAGAAACATTTTGAAGCAACA 549
    |||||
Db 421 TGGCTGTGATCAAAAATTTCAACAACAAATCAAACTTGAAGAAACATTTTGAAGCAACA 480
    |||||

OY 550 TGAATAACAACAAAACATATATATGTCAGTGTGAAGACTGTGAAGACCTTTAAGAA 609
    |||||
Db 481 TGAATAACAACAAAACATATATATGTCAGTGTGAAGACTGTGAAGACCTTTAAGAA 540
    |||||

OY 610 ACATCAGACAGTGAATAATCCATCAGTCCAGCATACCAATGAACCTTATTTCAAGTAC 669
    |||||
Db 541 ACATCAGACAGTGAATAATCCATCAGTCCAGCATACCAATGAACCTTATTTCAAGTAC 600
    |||||

OY 670 CCAAGAGAGATGTGGAAACACTTTCATACCCAGCAGCACTGAAAGAGATGCGCAAGGC 729
    |||||
Db 601 CCAAGAGAGATGTGGAAACACTTTCATACCCAGCAGCACTGAAAGAGATGCGCAAGGC 660
    |||||

OY 730 CCAAGAGAGTATGTATGTCAAAAAGAGATGTCCTTTGTGGCAAAAACATGAGAGAACT 789
    |||||
Db 661 CCAAGAGAGTATGTATGTCAAAAAGAGATGTCCTTTGTGGCAAAAACATGAGAGAACT 720
    |||||

OY 790 TCTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTGTGAGTATGCGCGGAAC 849
    |||||
Db 721 TCTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTGTGAGTATGCGCGGAAC 780
    |||||

OY 850 ATTTAAAGCAAGATTAACCTTAAGCAACATGAAACATCAT--GCCCGAAGAAAGGAT 907
    |||||
Db 781 ATTTAAAGCAAGATTAACCTTAAGCAACATGAAACATCATTTGCCGCAAGAAAGGAT 840
    |||||

OY 908 GTATGTGCG--TGTCAGAGAGAGGCTGTGG--AAGAACTATACTACTGT--GTTTATC 962
    |||||
Db 841 GTATGTGCGCGTGTGCCAAGAGGTTGTGAAAGAAACATATTCATCTGTGGGTATATCC 900
    |||||

OY 963 TCCAAAGCCATATC--CTCTCTTCATGAGCAAGCGCGCCCTTTT--GTGTGTAAC 1016
    |||||

```

Db 901 TCCCAAGCCATATCCCTCTCTTCCCTGAAGAAAGCCCGCTTTTGGGGGGGAGAC 960

OY 1017 ATGCTGGCTGTGGCAAAACA 1036

Db 961 ATGCTGGCTGTGGCGCAAAA 980

RESULT 15

BG830080 927 bp mRNA linear EST 22-MAY-2001

LOCUS 602764826F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4906710 5',

DEFINITION mRNA sequence.

ACCESSION BG830080

VERSION BG830080.1 GI:14177667

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 927)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM1810 row: d column: 07

High quality sequence stop: 880.

FEATURES

source

1..927

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4906710"

/clone_1lb="NIH_MGC_42"

/tissue_type="epitheloid carcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected for

average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-CDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC library. |"

BASE COUNT

264 a 244 c 229 g 190 t

ORIGIN

Query Match

Best Local Similarity 98.0%; Score 827.4; DB 12; Length 927;

Matches 908; Conservative 0; Mismatches 6; Indels 7; Gaps 7;

```

OY 98 GGGCGGGGGCGGGGTTCCCGGACAGTGTCTGGCAGCTGGCAGCGCGCTGCGG 157
    |||||
Db 1 GGGCGGGGGCGGGGTTCCCGGACAGTGTCTGGCAGCTGGCAGCGCGCTGCGG 60
    |||||

OY 158 CTGGAGAGCGCGCGCGCTGGATCGCGCGCGCGTGTGGCCGAGTGGTGGTCTTG 217
    |||||
Db 61 CTGGAGAGCGCGCGCGCTGGATCGCGCGCGCGTGTGGCCGAGTGGTGGTCTTG 120
    |||||

OY 218 ACCATCGCGCAGCGGTTTCATTTGACGCGCGGAGAGCTCACCTCGACCCCGCGCGCC 277
    |||||
Db 121 ACCATCGCGCAGCGGTTTCATTTGACGCGCGGAGAGCTCACCTCGACCCCGCGCGCC 180
    |||||

OY 278 GCGCTTCCAGAGAGTTCACTGTGCTCTTCCCTGACTGCAAGCGCAATTCACAGAAAGCC 337
    |||||
Db 181 GCGCTTCCAGAGAGTTCACTGTGCTCTTCCCTGACTGCAAGCGCAATTCACAGAAAGCC 240
    |||||

```

```
OY 338 TGGAGCTTGACGGCAGCTGTGCAAGCACACGGGGGAGAGACCATTGTTTGACTAT 397
|||||
Db 241 TGGAGCTTGACGGCAGCTGTGCAAGCACACGGGGGAGAGACCATTGTTTGACTAT 300
|||||
OY 398 GAAGGTGTGGCAAGGCTTCATCAGGACTACCATCTGAAGCCGCCACATTTGACTCAC 457
|||||
Db 301 GAAGGTGTGGCAAGGCTTCATCAGGACTACCATCTGAAGCCGCCACATTTGACTCAC 360
|||||
OY 458 ACAGGAGAAAAGCGTTGTTGTGCAAGCCACCTGGCTGATCAAAAATTCACACAAA 517
|||||
Db 361 ACAGGAGAAAAGCGTTGTTGTGCAAGCCACCTGGCTGATCAAAAATTCACACAAA 420
|||||
OY 518 TCAAACTTGAAGAAATTTTGAAGCAAAACATGAATAATCAACAAAATATATATGTC 577
|||||
Db 421 TCAAACTTGAAGAAATTTTGAAGCAAAACATGAATAATCAACAAAATATATATGTC 480
|||||
OY 578 AGTTTGAAGACTGTGAAGACCTTTAAGAAACATGACAGCTGAATAATCCATCAGTGC 637
|||||
Db 481 AGTTTGAAGACTGTGAAGACCTTTAAGAAACATGACAGCTGAATAATCCATCAGTGC 540
|||||
OY 638 CAGATACCAATGACCTCTATTCAAAGTACCCAGAGAGATGTGGAAACACTTTGCA 697
|||||
Db 541 CAGATACCAATGACCTCTATTCAAAGTACCCAGAGAGATGTGGAAACACTTTGCA 600
|||||
OY 698 TCAOCCAGCAAGCTGAACGACATGCCAAGGCCACGAGGCTATGTATGTCAAAAAGA 757
|||||
Db 601 TCAOCCAGCAAGCTGAACGACATGCCAAGGCCACGAGGCTATGTATGTCAAAAAGA 660
|||||
OY 758 TGTTCCTTTGTGGCAAAACATGAGCGAACTTCTGAACATGTGAGAGAAACCCAT-AA 816
|||||
Db 661 TGTTCCTTTGTGGCAAAACATGAGCGAACTTCTGAACATGTGAGAGAAACCCATAAA 720
|||||
OY 817 AGAGGAATACTATGTG-AACTATGCCGGAAGA-CATTTAAGCGAAGA-TTACTTTAA 873
|||||
Db 721 AGAGGAATACTATGTGAAAGTATGCCGAAMACATTAAACGCAAGATTACTTTAA 780
|||||
OY 874 GCAACACATGAAAACTCATGCCCCAGAAAGGATGTATGCCGTGTCCAAAGAAAGGCT- 932
|||||
Db 781 GCAACACATGAAAACTCATG-CCGAGAAAGGATCTATGTGCTGTCCAAAGAAAGGCTG 839
|||||
OY 933 GTGGAAGACCTATACTACTGTGTTTAAT-CTCCAAAGCCATATCTCTTCATGAG 991
|||||
Db 840 GTGGAAGACCTATACTACTGTGTTTAATCTCCAAAGGATATCTCTTCATGAG 899
|||||
OY 992 GAAAGCCGCCCTTTGTGTGT 1012
|||||
Db 900 GAAAGCCGCCCTTTGTGTGT 920
|||||
```

Search completed: February 10, 2003, 17:44:07
Job time : 2015.41 secs

THIS PAGE BLANK (USPTO)

0Y	1094	AAATGAAAGTC	AAAGCAAAAATTC	GTGAAAAACGAGTTT	GGCCCTC	ATCTAGT	1153
Db	199	AAAATGAAGTC	CAAAAGTC	CAAAAAATTC	GTGAAAAACGAGTTT	GGCCCTC	ATCTAGT 140
0Y	1154	GGATATATTC	CCCTCCCAAAACG	AAGCAAGGCAAGGCTT	ATCTTTGTG	CTCAAAACGGAAG	121.3
Db	139	GGATATATTC	CCCTCCCAAAAGCAAGG	CAAGGCTTATCTTTGTG	CTCAAAACGGAAG	80	
0Y	1214	TCACCCAACTGT	GTGGAAAGCAAGATGCT	CTGCACAGTTG	CAAGTCTACTAAC	CCCTGGCTAA	127.3
Db	79	TCACCCAACTGT	GTGGAAAGCAAGATGCT	CTGCACAGTTG	CAAGTCTACTAAC	CCCTGGCTAA 20	

```

RESULT 2
US-10-033-528-326/C
: Sequence 326: Application US/10033528
: Patent No. US2002013171A1
: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547C1
: CURRENT APPLICATION NUMBER: US/10/033,528
: CURRENT FILING DATE: 2001-12-26
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 336
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-033-528-326

```

Query Match	32.9%	Score 418.4	DB 12	Length 439
Best Local Similarity	99.8%	Pred. No. 8.4e-103		
Matches 419	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY 854	AAAGCCAAAGATTACCTTTAGCAACACATGAAAACTCATGCCCCGAGAAAGGATGTATGT			913
Db 439	AAAGCCAAAGATTACCTTTAGCAACACATGAAAACTCATGCCCCGAGAAAGGATGTATGT			380
QY 914	CGCTGCCAAGAGAGGGGTGTGGAAGACCTTACTACTGTGTTTAATCTCCAAAGCCAT			973
Db 379	CGCTGCCAAGAGAGGGGTGTGGAAGACCTTACTACTGTGTTTAATCTCCAAAGCCAT			320
QY 974	ATTCCTCTCTTCATGAGAGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAA			1033
Db 319	ATTCCTCTCTTCATGAGAGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAA			260
QY 1034	ACATTTGCAATGAACAACAAGTCTCACTAGCGATGCTGTTTACATGATCTTGACAAGAAG			1093
Db 259	ACATTTGCAATGAACAACAAGTCTCACTAGCGATGCTGTTTACATGATCTTGACAAGAAG			200
QY 1094	AAAAATGAAGCTCAAGAGTCAAAAAAATCTGTGAAAAAGGAGTTTGGCCCTCATCTCAGT			1153
Db 139	AAAAATGAAGCTCAAGAGTCAAAAAAATCTGTGAAAAAGGAGTTTGGCCCTCATCTCAGT			140
QY 1154	GGATATATTCCTCCCAAAAGAGAAACAAAGGCAAGGCTTATCTTTTGTGTCAAAAGGAGAG			1213
Db 139	GGATATATTCCTCCCAAAAGAGAAACAAAGGCAAGGCTTATCTTTTGTGTCAAAAGGAGAG			80
QY 1214	TCACCCCAACTGTGTGTGGAAGACAAGATGCTCTTCGACAGTTGCAAGTACTTACCTTGGCTTAA			1273
Db 79	TCACCCCAACTGTGTGTGGAAGACAAGATGCTCTTCGACAGTTCAGTACTTACCTTGGCTTAA			20

RESULT 3
US-09-920-300A-689
; Sequence 689, Application US/099203300A
; Patent No. US2002013672BA1
; GENERAL INFORMATION:

```

APPLICANT: King, Gordon E.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.547
CURRENT APPLICATION NUMBER: US/09/920.300A
CURRENT FILING DATE: 2001-07-31
NUMBER OF SEQ ID NOS: 1789
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 689
SEQ ID NO 689
LENGTH: 439
TYPE: DNA
ORGANISM: Homo sapiens
US-09-920-300A-689

Query Match          32.7%   Score 416.8; DB 10; Length 439;
Best Local Similarity 95.5%;   Pred. No. 2.3e-10;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Query Match	Best Local Similarity	Score	DB 10	Length
Matches 418	Conservative	99.58	Pred. No. 2.3e-102	Mismatches 2; Indels 0; Gaps
QY	854	AAACGCAAAAGATTACCTTAAGCAACACATGAAAAACTCATGCCCGAGAAAGGATGTATGT	913	
Db	1	AAACGCAAAAGATTACCTTAAGCAACACATGAAAAACTCATGCCCGAGAAAGGATGTATGT	60	
QY	914	CGCTGTCCAGAGAAAGCGTGTGGAAACACCTTATCACTGTGTTTAACTCCAAAGCCAT	973	
Db	61	CGCTGTCCAGAGAAAGCGTGTGGAAACACCTTATCACTGTGTTTAACTCCAAAGCCAT	120	
QY	974	ATCCCTCTCTTCATGAGAGAAAGCCGCCCTTTTGTGTGGAACATGCTGGCTGTGACAA	1033	
Db	121	ATCCCTCTCTTCATGAGAGAAAGCCGCCCTTTTGTGTGGAACATGCTGGCTGTGACAA	180	
QY	1034	ACATTTGCATGAAACAAAGTCTCACTAGCGCATGCTGTGTGTACATGATCTGTACAAAG	1093	
Db	181	ACATTTGCATGAAACAAAGTCTCACTAGCGCATGCTGTGTGTACATGATCTGTACAAAG	240	
QY	1094	AAATGAGGCTCAAAAGTCAAAATAATCTCGTGAATAAGGAGTTTGAGCCCTCATCTCAGT	1153	
Db	241	AAATGAGGCTCAAAAGTCAAAATAATCTCGTGAATAAGGAGTTTGAGCCCTCATCTCAGT	300	
QY	1154	GGATATATTCCTCCCAAAAGGAAACAAAGGAGCAAGGCTTATCTTTGTGTCAAAAACGAGAG	1213	
Db	301	GGATATATTCCTCCCAAAAGGAAACAAAGGAGCAAGGCTTATCTTTGTGTCAAAAACGAGAG	360	
QY	1214	TCACCCCAAGCTGTGGAGAACAGATGCTGTGACAGTGTGCACTACTTACCTTGGCTAA	1273	
Db	361	TCACCCCAAGCTGTGGAGAACAGATGCTGTGACAGTGTGCACTACTTACCTTGGCTAA	420	

```

RESULT 4
US-10-033-528-689
Sequence 689, Application US/10033528
Patent No. US20020131971A1
GENERAL INFORMATION:
APPLICANT: King, Gordon E.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.547C1
CURRENT APPLICATION NUMBER: US/10/033,528
CURRENT FILING DATE: 2001-12-26
NUMBER OF SEQ ID NOS: 1896
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 689
LENGTH: 439
TYPE: DNA
ORGANISM: Homo sapiens
US-10-033-528-689

```


Best Local Similarity	99.5%;	Pred. No.	2.3e-102;
Matches	418;	Conservative	0;
		Mismatches	2;
		Indels	0;
		Gaps	0;

QY	854	AAACGCAAACTTACCTTAAGCAACACATGAAACATCAATGCCCGAAGAAAGGATATGATG	913
Db	1	AAMCGCAAAATTTACCTTAAAGCAACATGAAAACTCATGCCCCAGAAAGGATTTATGT	60
QY	914	CGCTGTCCAAAGAGAGGCTGTGGAAAGACCTTACTACTGTGTTAAATCCCAAAAGCCAT	973
Db	61	CGCTGTCCAGAGAGAGGCTGTGGAAAGACCTTATACACTGTGTTAAATCTCCAAAGCCAT	120
QY	974	ATCCTCTCTTCCATGAGCAAGCGCCCTTTTGTGTGGAACATGCTGCTGTGGCAA	1033
Db	121	ATCCTCTCTTCATGAGCAAGGACCCCTTTTGTGTGGAACATGCTGCTGTGGCAA	180
QY	1034	ACATTTGCATGAAACAAACATCTCACTAGCAATGCTGTTGATCATGATCTGCACAAAG	1093
Db	181	ACATTTGCATGAAACAAACATCTCACTAGCAATGCTGTTGATCATGATCTGCACAAAG	240
QY	1094	AAATGAGCTCAAAGTCAAAAAATCTCGTGAAGAAAGGAGTTTGCCCTCATCTCACT	1153
Db	241	AAATGAGCTCAAAGTCAAAAAATCTCGTGAAGAAAGGAGTTTGCCCTCATCTCACT	300
QY	1154	GGATATATTCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCTTTGTGTCAAAACGAGAG	1213
Db	301	GGATATATTCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCTTTGTGTCAAAACGAGAG	360
QY	1214	TCACCCCAACTGTGTGGGAAGCAAGATGTCCTGCACAGTTGCAGTACTTACCTTGCGCTAA	1273
Db	361	TCACCCCAACTGTGTGGGAAGCAAGATGTCCTGCACAGTTGCAGTACTTACCTTGCGCTAA	420

RESULT 5
US-09-796-692-4064

```

Sequence 4064: Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Galger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY OF INFECTIONS
TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
PRIORITY FILING DATE: 2001-03-01
PRIORITY APPLICATION NUMBER: 60/186,126
PRIORITY FILING DATE: 2000-03-01
PRIORITY APPLICATION NUMBER: 60/190,479
PRIORITY FILING DATE: 2000-03-17
PRIORITY APPLICATION NUMBER: 60/200,545
PRIORITY FILING DATE: 2000-04-27
PRIORITY APPLICATION NUMBER: 60/200,303
PRIORITY FILING DATE: 2000-04-28
PRIORITY APPLICATION NUMBER: 60/200,779
PRIORITY FILING DATE: 2000-04-28
PRIORITY APPLICATION NUMBER: 60/200,999
PRIORITY FILING DATE: 2000-05-01
PRIORITY APPLICATION NUMBER: 60/202,084
PRIORITY FILING DATE: 2000-05-04
PRIORITY APPLICATION NUMBER: 60/206,201
PRIORITY FILING DATE: 2000-05-22
PRIORITY APPLICATION NUMBER: 60/218,950
PRIORITY FILING DATE: 2000-07-14
PRIORITY APPLICATION NUMBER: 60/222,903
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: 60/223,416
PRIORITY FILING DATE: 2000-08-04
PRIORITY APPLICATION NUMBER: 60/223,378
PRIORITY FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4064
LENGTH: 234

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4064

```

Query Match	18.4%;	Score 234;	DB 9;	Length 234;
Best Local Similarity	100.0%;	Pred. No. 1.9e-53;		
Matches 234;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	620	CTGAAATCCCATGCTGCTGCACCAATGCAATCTCTATTTCGAAGTGTACCACGGAAGA	679
Db	1	CTGAAATCCCATGCTGCTGCACCAATGCAATCTCTATTTCGAAGTGTACCACGGAAGA	60
QY	680	TGTGGGAACACATTGGCATCCCGACAGCTGAAAGACATGCGCAAGGCCACGAGGGC	739
Db	61	TGTGGGAACACATTGGCATCCCGACAGCTGAAAGACATGCGCAAGGCCACGAGGGC	120
QY	740	TATGATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATGAGCGGAACTTCTGAAACAT	799
Db	121	TATGATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATGAGCGGAACTTCTGAAACAT	180
QY	800	GTGAGAGAAACCCATAAAGAGGAATATCTATGTGAAGTATGCGGGAAACAATTT	853
Db	181	GTGAGAGAAACCCATAAAGAGGAATATCTATGTGAAGTATGCGGGAAACAATTT	234

RESULT 6

US-09-920-300A-610
; Sequence 610, Application US/09920300A
; Patent No. US20020136728A1

```

: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeline Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547
: CURRENT APPLICATION NUMBER: US/09/920,300A
: CURRENT FILING DATE: 2001-07-31
: NUMBER OF SEQ ID NOS: 1789
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 610
: LENGTH: 234
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-920-300A-610

```

RESULT 7

```
US-10-033-528-610
; Sequence 610, Application US/10033528
; Patent NO. US20020131971A1
; GENERAL INFORMATION:
```

```

: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547C1
: CURRENT APPLICATION NUMBER: US/10/033,528
: CURRENT FILING DATE: 2001-12-26
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 610
:
: LENGTH: 234
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
US-10-033-528-610

```

Query Match	18.4%	Score 234	DB 12	length 234
Best Local Similarity	100.0%	Pred. Nc.	1.9e-53	
Matches 234	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 620 CTGAAATATTCATCAGTGGCAGCATACCAATGAACCTATTTCAGTAGTACCACCGAAGA 679
|||||
Db 1 CTGAAATTCATCAGTGGCAGCATACCAATGAACCTATTTCAGTAGTACCACCGAAGA 60

680 TGTGGGAAACACTTTGCATCACCAGCAAGCTAAACACATGCCAAGGCC 739
 61 TGTGGGAAACACTTTGCATCACCAGCAAGCTAAACACATGCCAAGGCC 120

0Y TGTGTATGTCACAAAAGGATGTTCTCTTGTGGCAAAAACATGACGAACTTCGAAACAT 739
 121 TATGTATGTCACAAAAGGATGTTCTCTTGTGGCAAAAACATGACGAACTTCGAAACAT 180

QY	800	181	Db
GTGAGAGAAACCCATAAAGAGAGAAATACTATGTGAACATATGCGGAAAAACATTTT	853	GTGAGAGAAACCCATAAAGAGAGAAATACTATGTGAACATATGCGGAAAAACATTTT	234

RESULT 8
US-09-292-758-68/C
SEQUENCE 68 ANALYSIS OF US00000758

```

; GENERAL INFORMATION:
; APPLICANT: Burmer, Glenna C.
; ADDRESSEE: Bureau, Teresa D.

```

1 TITLE OF INVENTION: Nucleic Acid Sequences and Proteins
2
3 TITLE OF INVENTION: Associated With Aglyc
4
5 PTE REFERENCE: 017473-0011095

```

: CURRENT FILING DATE: 1999-04-14
:
: EARLIER APPLICATION NUMBER: US 60/081,887
:
: EARLIER FILING DATE: 1999-04-15

```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 363
```

! ORGANISM: Homo sapiens
US-09-292-758-68

Best Local Similarity 96.78; Pred. No. 3.5e-51;
Matches 263; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

Db 362 TGTGTGACATGCTGGCTGTGGCAAAACATTTCGAATGAAACAAGTCTCACTTAG 303

Db 302 GCATGCTGTTGACATGATCCTGCACAAGAAATGAAGCTCAAAAATCTC 243

Qy	1122	GTGAAAACGAGATTGGCCATCTCAGTGGATATATCCCTCCCAAAAGAAACAG	118
Db	242	GTGAAAACGAGATTGGCCATCTCAGTGGATATATCCCTCCCAAAAGAAACAG	183
Qy	1182	GGCAGGCTTATCTTTGTGCAAAACGAGATCACCCAACTGTGGAAAGCAAGATGC	124
Db	182	GGCAGGCTTATCTTTGTGCAAAACGAGATCACCCAACTGTGGAAAGCAAGATGC	123
Qy	1242	TCTCGACAGTTGCAGTACTTACCTTGGCTAA	1273
Db	122	TCTCGACAGTTGCAGTACTTACCTTGGCTAA	91

RESULT 9
US-09-995-973-2

```

; GENERAL INFORMATION:
; APPLICANT: CHOO, Yen

```

```

; FILE REFERENCE: 8325-2003 / G7-US1
; CURRENT APPLICATION NUMBER: US/09/995,973
CURRENT INVENTOR: 0000 0210

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
LENGTH: 847

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; COUND INFORMATION: Description of the feature (e.g. name, location, etc.)

```

Query Match	12.08	Score 152.4	DB 9	Length 947
Partial Similarity	17.08	Score 152.4	DB 9	Length 947

268 GCCGCGCCCGCGCTTCCAGAGATTATCTGCTTCCTTCCCTTGACTGCAGCGCCAATT 327

328 CAGCAAGCCTGGAAGCTTGACGGCCACCTGTGCAAGCACACGGGGAGAGACATTTGT 387

388 TTGTGA CTATGA AGCGTGTG CCAAGCCTT CATACAGGACTACCATCTGAGCGCCACAT 447

448 TCTGACTCAGACAGAGAAAGCCGTTTGTTCCTGCAGCCACTGGCTGATCAAAAAATT 507

508 CAACACAAATCAAACTTGAAGAAACATTTTGAACGCAAAACATGAAAAATCAACAAAAACA 567

568 ATATATATGCAGTTTGAAGACCTGAAGAAGACCTTAAGAACATCAGCAGCTGAAAT 627

628 CCATCAGTCCAGCATACCAATGAACCTCTATTCAAGTCTACCCAGGAAGATGTGGAA 687

688 ACACCTTGCATCACCACGACAGCTGAAACGACATGCCAACGCCCCACGAGGGCTA 741

RESULTS

```

Sequence 5, Application US/09732348
Patent No. US20020046419A1
GENERAL INFORMATION:
APPLICANT: Yen Choo, et al.
TITLE OF INVENTION: Regulated Gene Expression in Plants
FILE REFERENCE: 674538-2001
CURRENT APPLICATION NUMBER: US/09/732,348
CURRENT FILING DATE: 2000-12-07
NUMBER OF SEQ. ID NOS: 21
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 947
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc_feature
LOCATION: (723)..(908)
OTHER INFORMATION: transactivation domain of VP64, other features except c-myc tag (
NAME/KEY: misc_feature
LOCATION: (909)..(938)
OTHER INFORMATION: c-myc tag, other features except transactivation domain VP64 (115
US-09-732-348-5

```

[illegible]

```

RESULT 11
US-09-995-973-1
: Sequence 1, Application US/09995973
: Publication No. US20030024006A1
:
GENERAL INFORMATION:
:
: APPLICANT: CHOO, Iem
:
: APPLICANT: ULLMAN, Christopher G.
: TITLE OF INVENTION: GENE SWITCHES
: FILE REFERENCE: 8325-2003 / G-US1
: CURRENT APPLICATION NUMBER: US/09/995, 973
: CURRENT FILING DATE: 2002-03-19

```

```

: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 995
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:
: OTHER INFORMATION: TFIID/Zif-VP16
: US-09-995-973-1

```

Query Match	12.0%;	Score 152.4;	DB 9;	Length 995;
Best Local Similarity	57.6%;	Pred. No. 3.7e-31;		
Matches 273;	Conservative 0;	Mismatches 201;	Indels 0;	Gaps 0;

Oy	268	GC	GGGGGGGGGGCGCTCCGACAGAGGTTTCATCTGCTCTCTCCCTGACATCGACAGCAACCAATTA	327
Dd	26	GG	CGGTGGCGGTGGTGTATTAAGCGGGTACATCTGCTCTTTGGCGACATGGCGGGCTGCTTA	85
Oy	328	CAG	CAAAAGCCCTGGAAAGCTTTGACGGCGCACCTGTGTGCACACACAGGGGGAGAGACATTGTT	387
Dd	86	TAA	CAGAAACCTGGAAAGCTTCGACGGCGCATCTGTGCAAAACACACAGGACGAGAAACCATTTCC	145
Oy	368	TTTG	ACTATATGAAGGGTGTGGCAAGCCCTTCATCAGGAGCTTACCATTGAGCCGCCACAT	447
Dd	146	ATG	TAAAGGAAGAGATGTGGAAAGGCTTTTACCTCCTTCATCTACTTAAACCCGCCATC	205
Oy	448	TCTG	ACTACACAGAGAGAAAGCCGTTTGTGTGTGCAGCCACCGGGCTGANTCAAAAT	507
Dd	206	ACTG	CTACTACTGTCGGCGAGAAAACCTTCACATGTGTGACTCGGATGAGATGTGACTTTGAGATT	265
Oy	508	CAAC	CAAAATCAAACTTGAAGAAACATTTTGAACGCAAAACATGAAATAACACAAAACA	567
Dd	266	TACT	CAAAAGCCAAACATGAGAAAGCACTTTTAAACAGATTCCTCACTCAAGATCTGCGT	325
Oy	568	ATAT	TATGAGTTTGAAGCTGTAGGAAGACCTTTAAGAAACATGAGAGCGTGAAT	627
Dd	326	CTAT	TGTGGCAATTTTGGAACTGTGGCAAAAGCAATTCAAGAAACATCAATTAATTAAGGT	385
Oy	628	CCAT	CAGTCCAGCATACCAATGAACCTCTATTAAGTGTACCCAGGAAGAGATGTGGAA	687
Dd	386	TCAT	CAGTTGAGTGCACACAGCAGCTCCGCTATGCTTGCCCTGTGAGTCTGTGCATCG	445
Oy	688	ACACT	TTCGATCACCCAGCAAGCTGAAACGACATGCCAAGGCCACAGAGGCTA	741
Dd	446	CCGCT	TTTCTGGCTGGATGTAGCTTACCCTCCGCAATTCGCATCCACAGAGCCA	499

```

RESULT 12
US-09-732-348-4
: Sequence 4, Application US/09732348
: Patent No. US20020046419A1
: GENERAL INFORMATION:
: APPLICANT: ven Choo, et al.
: TITLE OR INVENTION: Regulated Gene Expression in Plants
: FILE REFERENCE: 674538-2001
: CURRENT APPLICATION NUMBER: US/09/732,348
: CURRENT FILING DATE: 2000-12-07
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 995
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (15)..(117)
: OTHER INFORMATION: translational Initiating ATG
: NAME/KEY: misc_feature
: LOCATION: (16)..(416)
: OTHER INFORMATION: Fingers 1 to 4 of TFIID
: NAME/KEY: misc_feature
: LOCATION: (308)..(416)

```



```
RESULT 14
US-09-833-381-1175
; Sequence 1175, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1175
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1111)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1175
```

Query Match 6.8%; Score 86.6; DB 10; Length 1111;

Best Local Similarity 61.1%; Pred. No. 1.9e-13;

Matches 140; Conservative 0; Mismatches 89; Indels 0; Gaps 0; *

```
QY 246 GCGAGAGCTAGCTCCGACCCCGCGCGCTTCCCGAGAGGTTCATCTGCTCCT 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 GCGTCCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCCAGAGCCACATCGCTCCT 63

QY 306 TCCTGACTGCGAGGCCAATTACGCAAGCGTGGAACTTGACGCCACCTGTGCAAGC 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 TCCGGAGCTGGCCAAAGCCCTACTACAAAGTCTCGACCTAAAGTCCGACCTCGGAGCG 123

QY 366 ACACGGGGGAGAGACCACTTTGTTGTGACTATGAAAGGTGGCAAGCCCTTCATCAGG 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 ACACAGGGGAGAGCGCTTTGCTGTGACTGGCGAGGCTCGACAGAAGTTCCGCCGCT 183

QY 426 ACTACACTGTAGCCGCCACATTCGACTCACACAGAGAAAGCCGCT 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 CCGACGAGCTGGCCCGCCACACCGGACGACGAGCGGCGAGAACGCTT 232
```

RESULT 15

US-09-920-300A-1747/c

; Sequence 1747, Application US/09920300A

; Patent No. US20020136728A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeline Joy

; APPLICANT: Xu, Jiangchun

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.547

; CURRENT APPLICATION NUMBER: US/09/920,300A

; CURRENT FILING DATE: 2001-07-31

; NUMBER OF SEQ ID NOS: 1789

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1747

; LENGTH: 368

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 223..323

; OTHER INFORMATION: n = A,T,C or G

```
Query Match 6.4%; Score 81; DB 10; Length 368;
Best Local Similarity 98.8%; Pred. No. 3.1e-12;
Matches 81; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 737 GCGTATGTATGTCAAAAGAGATGTTCTTTGTGGCAAAAACATGACGGAACCTTGAAA 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 GCGTATGTATGTCAAAAGAGATGTTCTTTGTGNCAAAAACATGACGGAACCTTGAAA 297

QY 797 CATGTGAGAAAACCCATAAG 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 CATGTGAGAAAACCCATAAG 275
```

Search completed: February 10, 2003, 12:58:32

Job time : 81.105 secs

THIS PAGE BLANK (USPTO)

D	b		120	ACGAGTCTCGGCACGTGGCAG6CCGCCCTTG6CCCMGG6CTTGGAG6GG6CG6CCG6CCCTCGGA	179
O	y	181	TCCGCGCGGCGCTGGTGGCCGCGAATGCGGTGTCCTTTGACCATTGCGCGAGCGCTTACATTCG	240	
D	b	180	TCCGCCCGGCGCTGGTGGCCGAGTGGGTGTCCTTTGACCATTGCGCGAGCGCTTACATTCG	239	
O	y	241	AGCGGSGGAGAGCTCACCTCGAGCCCCGGGCGCGCCGCTTCCAGAGAGTCAATCTGG	300	
D	b	240	AGCGGSGGAGAGCTCACCTCGAGCCCCGGGCGCGCCGCTTCCAGAGAGTCAATCTGG	299	
O	y	301	CTTCCTTCCCCTACTGACAGCGCCAAATTAACAGCAAAGCCTGGAGACTTGAACGCGACCTGTG	360	
D	b	300	CTTCCTTCCCCTACTGACAGCGCCAAATTAACAGCAAAGCCTGGAGACTTGAACGCGACCTGTG	359	
O	y	361	CAAGCACACGCGGAGAGACCATTTGTTTGTGACTATGAAGGGTGTGGCAGGCGCTTCAT	420	
D	b	360	CAAGCACACGCGGAGAGACCATTTGTTTGTGACTATGAAGGGTGTGGCAGGCGCTTCAT	419	
O	y	421	CAGGAGCTACCATTTGAGCGCGCCACATTTGACTCACACAGGAGAAAAGCCGTTTGTGG	480	
D	b	420	CAGGAGCTACCATTTGAGCGCGCCACATTTGACTCACACAGGAGAAAAGCCGTTTGTGG	479	
O	y	481	TGCAGCCACTGGCTGTGATCAAAAAATTCACACAAAAATCAACTGTGAAGAAACATTTGA	540	
D	b	480	TGCAGCCACTGGCTGTGATCAAAAAATTCACACAAAAATCAACTGTGAAGAAACATTTGA	539	
O	y	541	ACGCAAAACATGAAAAATCAACAAAAACAATATATATGCACTGTTTGAAGACTGTAAAGAAGAC	600	
D	b	540	ACGCAAAACATGAAAAATCAACAAAAACAATATATGCACTGTTTGAAGACTGTAAAGAAGAC	599	
O	y	601	CTTTAAGAAACATCAGCAGCTGAAAATTCATCAGTGGCAGCATPCCAATGACCTCTATT	660	
D	b	600	CTTTAAGAAACATCAGCAGCTGAAAATTCATCAGTGGCAGCATPCCAATGACCTCTATT	659	
O	y	661	CAAGTGTACCCGAGGAAGATGTGGAAAACCTTGTGCATCCAGCAGAGCTGAAGAGACA	720	
D	b	660	CAAGTGTACCCGAGGAAGATGTGGAAAACCTTGTGCATCCAGCAGAGCTGAAGAGACA	719	
O	y	721	TGCCAAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG	780	
D	b	720	TGCCAAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG	779	
O	y	781	GACGSAACCTTGTGAACATGTGAGAGAAACCCATTAAGAGGAAATACTATGTCAAGTATG	840	
D	b	780	GACGSAACCTTGTGAACATGTGAGAGAAACCCATTAAGAGGAAATACTATGTCAAGTATG	839	
O	y	841	CCGGA AAA CATTTTAAGCGAAAGATPACCTTAAGCAACATGAAAAATCATGCCCCACGA	900	
D	b	840	CCGGA AAA CATTTTAAGCGAAAGATPACCTTAAGCAACATGAAAAATCATGCCCCACGA	899	
O	y	901	AAGGATGTATGTGCGTGTCCAAAGAGAAGCGTGTGANAAGAACTATACTACTGTGTTTAA	960	
D	b	900	AAGGATGTATGTGCGTGTCCAAAGAGAAGCGTGTGANAAGAACTATACTACTGTGTTTAA	959	
O	y	961	TCTCCAAAGCCATATCTCTCTCTTCCATGAGGAAAAGCCGCGCTTTGTGTGAACATGC	1020	
D	b	960	TCTCCAAAGCCATATCTCTCTCTTCCATGAGGAAAAGCCGCGCTTTGTGTGAACATGC	1011	
O	y	1021	TGGGCTGTGGCAAAACATTTTGGCAAGAAACAAGCTCAGTAAGCAATGCTGTGTACATGA	1080	
D	b	1020	TGGGCTGTGGCAAAACATTTTGGCAAGAAACAAGCTCAGTAAGCAATGCTGTGTACATGA	107	
O	y	1081	TCTGTGACAAGAGAAATGAAGCTGCAAGATCAAAAAATCTGTGAAAANAAC - GGAAGTTTGG	113	
D	b	1080	TCTGTGACAAGAGAAATGAAGCTGCAAGATCAAAAAATCTGTGAAAANAAC - GGAAGTTTGG	113	
O	y	1140	CCTCTCATCTCAGTGAATATATCTCTTCCCAAAAGGAAACAAGGCAAGCTTATCTTTGT	119	
D	b	1140	CCTCTCATCTCAGTGAATATAT - CCTCTCCAAAAGGAAACAAGGCAAGGCTTATCTTTGT	119	
O	y	1200	GTCAAAACGGAGAGTCAACCCCAACGTGTGTGGAAGACAAGATGCTGTGACAGTTGCAAGTAC	125	
D	b	1139	GTCAAAACGGAGAGTCAACCCCAACGTGTGTGGAAGACAAGATGCTGTGACAGTTGCAAGTAC	125	

QY	1260	TTACCCCTGGC	1270
Db	1259	TTACCCCTGGC	1269

RESULT 3
HS-09-492-985-11

```

1 Sequence 11, Application US/09/492985
2 Patent No. 6376240
3
4 GENERAL INFORMATION:
5
6 APPLICANT: Song, An M.
7 APPLICANT: Chen, Ya-Fen
8 APPLICANT: Krensky, Alan M.
9 TITLE OF INVENTION: RFLAT-1: A Transcribed
10 TITLE OF INVENTION: Activates RANRES Gene
11 FILE REFERENCE: SUN-113P
12 CURRENT APPLICATION NUMBER: US/09/492,985
13 CURRENT FILING DATE: 2000-01-27
14 EARLIER APPLICATION NUMBER: 60/111,576
15 EARLIER FILING DATE: 1999-01-27
16 NUMBER OF SEQ ID NOS: 11
17
18 SOFTWARE: SEQID for Windows Version 4.0
19
20 SEQ ID NO 11
21
22 LENGTH: 1471
23
24 TYPE: DNA
25 ORGANISM: mouse
26
27 US-09-492-985-11

```

Query Match

Matches 17

```
5.08; Score 63.2; DB 4; Length 1471;
48.98; Pred. No. 4.2e-07;
0; Mismatches 178; Indels 0; Gaps 0;
```

QY	127	CTGGCAGCTGGACACGGCCCTGGCCCTTGGAGAGCCGGCCCTGGATCCGC	186
Db	715	CCCGGGACACAGAGGCGCGCGCGCGCCGCCCCCAAGCCCGCCCTGGAGCGCAGCCCAAGGC	774
QY	187	GGCCGTGCTGCCGAGTGGGTGTGCTCTTACCATGCGCAGCCGTTTCATTGGACCGCG	246
Db	775	GGCATTTTGAGACGAGAACCCGGCCCGCGGGAGGCGGAGGCTTGCCCTGACACAAAGGG	834
QY	247	CGAGAGCTACAGCTCGACCGCCGCGCGCGCGCTTCCAGAGAGTTCACTTCGCTT	306
Db	835	TGGGGGAGGCCGAGACCGCGCGGACCTCGAATCCCCCGAGAGAGACACAAGTCCACTA	894
QY	307	CCCTACTGCAAGCCCAATTACAGCAAAAGCTGGAAGCTTGACGCGCACTGTGACAGA	366
Db	895	CCCGGGCTGCGGAGAAAGTTTACGGGAAATCTTCCACCTCAAGGCGCGCACTGGAACATCA	954
QY	367	CACGGGGGAGACCACTTTGTTTGACTATGAAAGGCTGTGCAAGGCTTTCATCAGGA	426
Db	955	CACAGGTGAGAGGCTTTCCGCTGAGCTGGCAGAGATGCACACAAGTTTCGACGCTC	101
QY	427	CTACCATCTGAACCGCCACACATTTGACTACACACAGGAGAAAGCCGTT	474
Db	1015	GGACGACGCTGGACGCGCACTTCCGACCCACGCGGGAGAAAGTT	1062

RESULT 4

```

1 Sequence 1, Application US/09492385
2 Patent No. 6376240
3 GENERAL INFORMATION:
4 APPLICANT: Song, An M.
5 APPLICANT: Chen, Ya-Fen
6 APPLICANT: Krensky, Alan M.
7 TITLE OF INVENTION: RELAT-1: A Transcription Factor That
8 TITLE OF INVENTION: Activates RANTES Gene Expression
9 FILE REFERENCE: SUN-113P
10 CURRENT APPLICATION NUMBER: US/09/492,985
11 CURRENT FILING DATE: 2000-01-27
12 EARLIER APPLICATION NUMBER: 60/117,576
13 EARLIER FILING DATE: 1999-01-27

```



```

; APPLICANT: Spelsberg, T. C.
; TITLE OF INVENTION: DNA ENCODING TGF-BETA INDUCIBLE
; TITLE OF INVENTION: EARLY FACTOR-1 (TIEF-1), A GENE EXPRESSED
; TITLE OF INVENTION: BY OSTEOBLASTS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,227A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.157US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-359-3260
; TELEFAX: 612-359-3263
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-570-227A-1

Query Match      4.2%  Score 53.2; DB 2; Length 2881;
Best Local Similarity 54.6%; Pred. No. 0.00023;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 287 AGGAGGTCATCTCTCTCTCCCTGACTGCAGCGCCCAATTACAGCAAGCCTGGAGCTT 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1185 AGGAGTCACATCTGTAGCCACCGAGGATGTGGCAAGACATCTTAAAGTTCCCATCTG 1244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 GACCGCACCTGTGCAGACGAGGAGAGACCATTTGTTGTGACTATGAAGGGTGT 406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1245 AAGGCCACACGAGAGACGACAGAGAGAGAAAGCCTTACACTGTAGTGGAAAGTTGT 1304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 GGCAAGGCTTCATCAGAGGACTACCATCTGAGCGCCACATTCGTGACTCAGAGAGAA 466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1305 GAAAGGAGGTTTGGCCGCTTGTGATGAAGTGTCCAGACACAGGCAAGCCACAGGGTGTAG 1364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 467 AAGCGTTTGTG 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1365 AAGAAATTTGGGTG 1378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-09-077-991-1
; Sequence 1, Application US/09077991
; Patent No. 6207375
; GENERAL INFORMATION:
; APPLICANT: Subramaniam, M.
; APPLICANT: Spelsberg, T.C.
; APPLICANT: Roche, P.C.
; TITLE OF INVENTION: TGF-Beta inducible early factor-1
; TITLE OF INVENTION: (TIEF-1) and a method to detect breast cancer
; FILE REFERENCE: 150.157US2

```

```

; CURRENT APPLICATION NUMBER: US/09/077,991
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: PCT/US96/19555
; EARLIER FILING DATE: 1996-12-11
; EARLIER APPLICATION NUMBER: US 08/570,227
; EARLIER FILING DATE: 1995-12-11
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2881
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-077-991-1

Query Match      4.2%  Score 53.2; DB 4; Length 2881;
Best Local Similarity 54.6%; Pred. No. 0.00023;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 287 AGGAGTCATCTCTCTCTCCCTGACTGCAGCGCCCAATTACAGCAAGCCTGGAGCTT 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1185 AGGAGTCACATCTGTAGCCACCGAGGATGTGGCAAGACATCTTAAAGTTCCCATCTG 1244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 GACCGCACCTGTGCAGACGAGGAGAGACCATTTGTTGTGACTATGAAGGGTGT 406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1245 AAGGCCACACGAGAGACGACAGAGAGAGAAAGCCTTACACTGTAGTGGAAAGTTGT 1304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 GGCAAGGCTTCATCAGAGGACTACCATCTGAGCGCCACATTCGTGACTCAGAGAGAA 466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1305 GAAAGGAGGTTTGGCCGCTTGTGATGAAGTGTCCAGACACAGGCAAGCCACAGGGTGTAG 1364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 467 AAGCGTTTGTG 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1365 AAGAAATTTGGGTG 1378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESS: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109

```



```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 05/08/946,241B
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
OS-08-946-241B-8

```

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 05/08/946,241B
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creasod, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
US-08-946-241B-8

```

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 05/08/946,241B
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creasod, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
OS-08-946-241B-8

```


22	774.6	60.8	802	17	US-09-040-623-24482	Sequence 24482, A
23	774.6	60.8	802	33	US-09-898-8888-24482	Sequence 24482, A
24	774.6	60.8	802	33	US-09-898-8888-24482	Sequence 24482, A
25	613	48.6	776	25	US-09-652-814-6814	Sequence 6814, A
26	555.4	43.2	557	20	US-09-535-897-7605	Sequence 7605, A
27	565.4	42.9	1359	25	US-09-644-873-10092	Sequence 10092, A
28	546.4	42.9	1359	27	US-09-699-998-9545	Sequence 9545, A
29	541.2	42.9	1467	80	US-00-360-207-7415	Sequence 7415, A
30	495	38.9	690	1	PCT-US01-18569-803	Sequence 803, App
31	495	38.9	690	43	US-09-644-873-7308	Sequence 803, App
32	470.8	37.0	965	25	US-09-306-609-4878	Sequence 7308, App
33	459.2	36.9	497	17	US-09-606-809-4878	Sequence 4878, App
34	459.2	36.9	497	19	US-09-522-251-4878	Sequence 4878, App
35	469.2	36.9	497	38	US-10-025-600-4878	Sequence 4878, App
36	453.4	35.6	618	20	US-09-535-897-7597	Sequence 7597, App
37	429.2	33.7	650	81	US-60-377-240-6706	Sequence 6706, App
38	418.4	32.9	439	34	US-09-820-300-326	Sequence 326, App
39	418.4	32.9	439	34	US-09-820-300A-326	Sequence 326, App
40	418.4	32.9	439	38	US-10-933-528-326	Sequence 326, App
41	418.4	32.9	439	39	US-10-099-926-326	Sequence 326, App
42	416.8	32.7	439	34	US-09-920-300-689	Sequence 689, App
43	416.8	32.7	439	34	US-09-920-300A-689	Sequence 689, App
44	416.8	32.7	439	38	US-10-033-528-689	Sequence 689, App
45	416.8	32.7	439	39	US-10-039-926-689	Sequence 689, App

ALIGNMENTS

```

RESULT 1
US-09-831-426-1
: Sequence 1, Application US/09631426
: GENERAL INFORMATION:
: APPLICANT: Hoechst Marion Roussel
: APPLICANT: Bordon-Pallier, F.
: APPLICANT: Roche, C.
: TITLE OF INVENTION: Human htfIIa gene and coded htfIIa protein
: FILE REFERENCE: 146,164
: CURRENT APPLICATION NUMBER: US/09/831,426
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: Patentin Vers. 2.0
: SEQ ID NO 1
: LENGTH: 1273
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (176)..(1270)
: US-09-831-426-1

```

Query Match	100.0%	Score 1273;	DB 32;	Length 1273;
Best Local Similarity	100.0%	Pred. No. 1.9e-279;		
Matches 1273; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	ATGGCGACACACGGGCGCCGAGAGCGGGGCGGGCTGTGTGACACCGGGCGGCTCCCGGAAT	60
Dp	1	ATGGCGACACACGGGCGCCGAGAGCGGGGCGGGCTGTGTGACACCGGGCGGCTCCCGGAAT	60
QY	61	GTGCGCGCGGTGCGCCGAAAGTTACAGCAGGAGCGCGTGGCGCGCGGCGGTTCGCGGC	120
Dp	61	GTGCGCGCGGTGCGCCGAAAGTTACAGCAGGAGCGCGTGGCGCGCGGCGGTTCGCGGC	120
QY	121	ACGTCTCTCGGCACGTGGGACCGCGCCCTGTGGCTCTTGGAGAGCGCCCGCCCTCTGA	180
Dp	121	ACGTCTCTCGGCACGTGGGACCGCGCCCTGTGGCTCTTGGAGAGCGCCCGCCCTCTGA	180
QY	181	TCCGCCCGGCGGTGTGCGCCGAGTGGGTGTGCTTACCATGCGCGACGGTTCATATTC	240
Dp	181	TCCGCCCGGCGGTGTGCGCCGAGTGGGTGTGCTTACCATGCGCGACGGTTCATATTC	240
QY	241	AGCGCGGAGACGCTCAGCTCGGACCCCGCGCGCCCGCGCTTCCAGGAGTTCAATCTG	300
Dp	241	AGCGCGGAGACGCTCAGCTCGGACCCCGCGCGCCCGCGCTTCCAGGAGTTCAATCTG	300

Db	241	AGGGGGAGAGCTCABGTCGACCCCGCGCGCCCGCGCTTCCAGAGAGGTTGATCTG	300
Qy	301	CTCTCTTCCTGACTGCGAGGGCCAAATTACGACAAAGCTGGAAGCTTGAACGGCACCTGTG	360
Db	301	CTCTCTTCCTGACTGCGAGGGCCAAATTACGACAAAGCTGGAAGCTTGAACGGCACCTGTG	360
Qy	361	CAACACACACGGGGAGAGACAAATTGTTTGTGACTATGAAGGGTGTGGCAAGGCGCTTCAT	420
Db	361	CAACACACACGGGGAGAGACAAATTGTTTGTGACTATGAAGGGTGTGGCAAGGCGCTTCAT	420
Qy	421	CAGGAGCTACCATCTGAGCCGCCACATTCGTGACTCACAGAGAGAAAGCCGTTGTGTTG	480
Db	421	CAGGAGCTACCATCTGAGCCGCCACATTCGTGACTCACAGAGAGAAAGCCGTTGTGTTG	480
Qy	481	TGCGAGCCAGTGGGTGTGATCTCAAAAATTCAACCAAAATCAAACTGAGAGAAACATTTTGA	540
Db	481	TGCGAGCCAGTGGGTGTGATCTCAAAAATTCAACCAAAATCAAACTGAGAGAAACATTTTGA	540
Qy	541	ACGGAAACATGAANAATCAACAAAACMAATATATATGCACTGTTTGAAGACTGTAAAGAC	600
Db	541	ACGGAAACATGAANAATCAACAAAACMAATATATATGCACTGTTTGAAGACTGTAAAGAC	600
Qy	601	CTTTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCACAGATACCAATGAACCTCTATT	660
Db	601	CTTTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCACAGATACCAATGAACCTCTATT	660
Qy	661	CAAGTATACCCAGAGAAAGATGTGGGAACACTTTCGATCACCACAGCTGAAGACGCA	720
Db	661	CAAGTATACCCAGAGAAAGATGTGGGAACACTTTCGATCACCACAGCTGAAGACGCA	720
Qy	721	TGCGAAGGCCACGAGGGCTATGTATGTCAAAAAAGGATGTTCTTTGTGGCAAAAAACATG	780
Db	721	TGCGAAGGCCACGAGGGCTATGTATGTCAAAAAAGGATGTTCTTTGTGGCAAAAAACATG	780
Qy	781	GACGGAACCTCTGAACATGTGAGAAACCCATAAAGAGAAATACATATGTGAAGTATG	840
Db	781	GACGGAACCTCTGAACATGTGAGAAACCCATAAAGAGAAATACATATGTGAAGTATG	840
Qy	841	CCGGAAACACTTTAAACGCAAAAGATTACCTTAAAGCAACATGAANAACATCAGCCCCAGA	900
Db	841	CCGGAAACACTTTAAACGCAAAAGATTACCTTAAAGCAACATGAANAACATCAGCCCCAGA	900
Qy	901	AAGGATGTATGTCTGCTGTCCAAGAGAGAGGCTGTGAGAGACCTTACTACTGTGTTTAA	960
Db	901	AAGGATGTATGTCTGCTGTCCAAGAGAGAGGCTGTGAGAGACCTTACTACTGTGTTTAA	960
Qy	961	TCTCAAAAGCCATATCTCTCTCTCTCCATATGAGAAAGCCGCCCTTTTGTGTGTGAACATGC	1020
Db	961	TCTCAAAAGCCATATCTCTCTCTCTCCATATGAGAAAGCCGCCCTTTTGTGTGTGAACATGC	1020
Qy	1021	TGCGTGTGGCAAAACATTTGCAATGGAACAAAGCTCACTAGGACATGCTGTTTACATGA	1080
Db	1021	TGCGTGTGGCAAAACATTTGCAATGGAACAAAGCTCACTAGGACATGCTGTTTACATGA	1080
Qy	1081	TCTGACACAAAGAAATGAAGCTCAAAGTCAAAAAAATCTGTGAAAAAGGAGTTTGGC	1140
Db	1081	TCTGACACAAAGAAATGAAGCTCAAAGTCAAAAAAATCTGTGAAAAAGGAGTTTGGC	1140
Qy	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAAAGGAGGCAAGGCTTATCTTTGTG	1200
Db	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAAAGGAGGCAAGGCTTATCTTTGTG	1200
Qy	1201	TCAAAACGAGAGTCAACCCAACTGTGTGGAAGCAAGATGCTCTCGACAGTTGCGAGTACT	1260
Db	1201	TCAAAACGAGAGTCAACCCAACTGTGTGGAAGCAAGATGCTCTCGACAGTTGCGAGTACT	1260
Qy	1261	TACCTTGGCTAA 1273	
Db	1261	TACCTTGGCTAA 1273	

RESULT 2
US-09-831-426-3

Db	121	ACGTCGTGGGCAAC	CTGTGGCAAGCCGCGCTGGCCCTGGGCTTGGAGAGGCGCGCGCCCTG	180
Qy	181	TCGCCCGCCGGTGGTGGCGGATGGGTGTCTCTTTGACCATGGCGGCTTCATTTGC	240	
Db	181	TCGCCCGCCGGTGGTGGCGGATGGGTGTCTCTTTGACCATGGCGGCGCTTCATTTGC	240	
Qy	241	AGCGGGGACAGCTCAGCTCCGAGCCCGCGCGCGCGCTTCCGAGAGTTCACTG	300	
Db	241	AGCGGGGACAGCTCAGCTCCGAGCCCGCGCGCGCGCTTCCGAGAGTTCACTG	300	
Qy	301	CTCTCTCCCTGACTGCTAGCGCCCAATTACAGAAAGGCTTGAAGCTTGAAGCGGCACTGTG	360	
Db	301	CTCTCTCCCTGACTGCTAGCGCCCAATTACAGAAAGGCTTGAAGCTTGAAGCGGCACTGTG	360	
Qy	361	CAAGCACACGGGGGAGAGACATTTGTTTGTGACTATGAAGGGTGTGGCAAGGCGCTTCAT	420	
Db	361	CAAGCACACGGGGGAGAGACATTTGTTTGTGACTATGAAGGGTGTGGCAAGGCGCTTCAT	420	
Qy	421	CAGGACTACCATCTGTAGCGCCGACATTTGACTCACACAGAGAAAGCCGTTTGTG	480	
Db	421	CAGGACTACCATCTGTAGCGCCGACATTTGACTCACACAGAGAAAGCCGTTTGTG	480	
Qy	481	TGCAGCCACTGGGCTGTATCAAAAATTCAACAAATCAAACTTGAAGAAACATTTTGA	540	
Db	481	TGCAGCCACTGGGCTGTATCAAAAATTCAACAAATCAAACTTGAAGAAACATTTTGA	540	
Qy	541	ACGGAAACATGAAAAATCACAAAAACAAATATATGTCAGATTTTGAAGACTGTAGAAAGAC	600	
Db	541	ACGGAAACATGAAAAATCACAAAAACAAATATATGTCAGATTTTGAAGACTGTAGAAAGAC	600	
Qy	601	CTTTAAGAAACATCAGCAGCTGAAATCCATCAGTCCAGCATACCATGAACTCTATT	660	
Db	601	CTTTAAGAAACATCAGCAGCTGAAATCCATCAGTCCAGCATACCATGAACTCTATT	660	
Qy	661	CAAGGTATACCAGGAAGATGTGGGAAACACTTGGATACCCACAGCAAGCTGAAGACGCA	720	
Db	661	CAAGGTATACCAGGAAGATGTGGGAAACACTTGGATACCCACAGCAAGCTGAAGACGCA	720	
Qy	721	TGGCAAGGCCACGAGGGCTATGTATGTCAAAAAGGATGTTCTTTGTGGCAAAACATG	780	
Db	721	TGGCAAGGCCACGAGGGCTATGTATGTCAAAAAGGATGTTCTTTGTGGCAAAACATG	780	
Qy	781	GACGGAACTTCTGAAACATGTGAGAGAAACCCTTAAGAGAAATACATGTGAAGTATG	840	
Db	781	GACGGAACTTCTGAAACATGTGAGAGAAACCCTTAAGAGAAATACATGTGAAGTATG	840	
Qy	841	CCGGAAAAATTTTAAAGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGCCCCGGA	900	
Db	841	CCGGAAAAATTTTAAAGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGCCCCGGA	900	
Qy	901	AAGGATGTATGTGCTGCTCAAGAGAGAGGCTGTGAGAAACCTATACTGTTTAA	960	
Db	901	AAGGATGTATGTGCTGCTCAAGAGAGAGGCTGTGAGAAACCTATACTGTTTAA	960	
Qy	961	TCTCCAAAGCATATCCTCTCTCTTCCATGAGAGAAAGCCGCTTTTGTGTGAACTATG	1020	
Db	961	TCTCCAAAGCATATCCTCTCTCTTCCATGAGAGAAAGCCGCTTTTGTGTGAACTATG	1020	
Qy	1021	TGGCTGTGGCAAAACATTTTGCATGAAACAAAGTCTCAGTAGCAATGGCTGTACATGA	1080	
Db	1021	TGGCTGTGGCAAAACATTTTGCATGAAACAAAGTCTCAGTAGCAATGGCTGTACATGA	1080	
Qy	1081	TCCTGACAAAGAAATGAAAGCTCAAAAGTCAAAAAATCTGTAAGAAAAAGGAGCTTGGC	1140	
Db	1081	TCCTGACAAAGAAATGAAAGCTCAAAAGTCAAAAAATCTGTAAGAAAAAGGAGCTTGGC	1140	
Qy	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGAAAGGCAAGGCTTATCTTTGTG	1200	
Db	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGAAAGGCAAGGCTTATCTTTGTG	1200	
Qy	1201	TCAAAACGGAGAGTCAACCAACTGTGTGGAAGCAAGATGCTCTGCACAGTTGCAGTACT	1260	
Db	1201	TCAAAACGGAGAGTCAACCAACTGTGTGGAAGCAAGATGCTCTGCACAGTTGCAGTACT	1260	

```

Oy      1261  TACCGTTGGCTAA  1273
        |||||||
Db      1261  TACCGTTGGCTAA  1273

RESULT 4
US-09-831-426C-3
: Sequence 3, Application US/09831426C
: GENERAL INFORMATION:
: APPLICANT: Hoechst Marion Roussel
: APPLICANT: Borden-Pallier, F.
: APPLICANT: Rocher, C.
: TITLE OF INVENTION: Human htfIIIA gene and coded htfIIIA protein
: FILE REFERENCE: 146.1364
: CURRENT APPLICATION NUMBER: US/09/831.426C
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Vers. 2.0
: SEQ ID NO 3
: LENGTH: 1273
: TYPE: DNA
: ORGANISM: Human
US-09-831-426C-3

```

```
Query Match      100.0%;  Score 1273;  DB 32;  Length 1273;
```

Best Local Similarity 100.0%; Pred. NO. 1.3e-2/9;
Matches 1273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGCGCAGCAGCGCGCGCGGACCGCGGGCGGTGCTGTGACCGCGCGCGCTCCCGGAAGT 60

1 ATGCGCAGCCGCCGCCTGGACCGCGCGTTCGGAAGT 60

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible][illegible][illegible][illegible]

181 TCCGCCGCGCTGGTCGCCGAGTCCGTGTCTGCTCTTGAACCATCCGCCGACCGCTTCAATTGC 240

181 TCCGCCGCCGTGGTCGCCGAGTCGGTGTCTCCTTGACCATCGCCGACGGCTTCATTGC 240

241 AGCCGGAGAGCTCAGCTCCGACCCCGCCGCCGCTTCCAGAGGTTCACTG 300

241 AGCCGGGAGAGCTCAGCTCCGACCCCGCCGCCCCCGCTTCCAGAGGTTCACTG 300

301 CTCCTTCCCTGACTGCAGCGCCAATTACAGCAAGCCTGGAAGCTTGACGGGCACCTGTG 360

301 CTCTTCCCTGACTGCAGCGCCAATTACAGCAAAGCCTGGAAGCTTGACGGCACCCTGTG 360

361 CAAGCACACGGGGGAGAGACCATTGTTGTGACTATGAAGGTTGTGGCAAGCCTTCAT 420

361 CAAGCACCAGGGGAGAGACCATTCTTTCTGCACTATGAGGGTGTGGCAAGGCTTCAT 420

A21 CCCCCCCTCCCTCCCCCCCCCCCCCCCCCCCTTTTTC 490

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

481 1GCA GCCAC1GGC1G1GAC1CAAAA11CAAC1CACAC1CAAAA1CAAA11GAAAGAAAC1111GA 240

481 TGCAGCCACTGGCTGTGATCAAAAATTCACACCAAAATCAACTTGAAGAACAATTTTGA 540

541 ACGCAACATGAAATCAACAAAAACAATATATATGCGAGTTTGAAGACTGTAAGAAGAC 600

541 ACGCAACATGAAATCAACAAAACAATATATATGCACTTTGAAGACTGTAAGAAGAC 600

601 CTTAAGAACATCAGCAGCTGAAATCCATCAGTGCAGCATACCAATGAACCTCTATT 660

601 CTTAAGAACATCAGCAGCTGAAATCCATCAGTGCCAGCATACCAATGAACCTCTATT 660

Qy	661	CAAGTCATACCCAGGAAGATGTCGGAAACACTTTGCTATCCACGACAGCTGAAACGACA	720
Db	661	CAAGTCATACCCAGGAAGATGTCGGAAACACTTTGCTATCCACGACAGCTGAAACGACA	720
Qy	721	TGCCAAGGCCACAGAGGGCTATGTATGTCAAAAAGATGTTCCTTTGGCGAAAAACATG	780
Db	721	TGCCAAGGCCACAGAGGGCTATGTATGTCAAAAAGATGTTCCTTTGGCGAAAAACATG	780
Qy	781	GACGGAACCTTCTGAACCATGTGAGAGAAACCATAAGAGAAATACTATGTGAAGTATG	840
Db	781	GACGGAACCTTCTGAACCATGTGAGAGAAACCATAAGAGAAATACTATGTGAAGTATG	840
Qy	841	CCGGAANAACATTTAAACGCAAGATTAACCTTAAGCAACACATGAAAACTCATGCCCCAGA	900
Db	841	CCGGAANAACATTTAAACGCAAGATTAACCTTAAGCAACACATGAAAACTCATGCCCCAGA	900
Qy	901	AAGGATGTATGTGCGCTGTCGAAGACAGACGTGTGGAAGAACCTTACTACTGTGTTTAA	960
Db	901	AAGGATGTATGTGCGCTGTCGAAGACAGACGTGTGGAAGAACCTTACTACTGTGTTTAA	960
Qy	961	TCTCCAAAGCCATATTCCTCTCCTTCCATGAGGAAGCCGCCCTTTTGTGTGGAACATGC	1020
Db	961	TCTCCAAAGCCATATTCCTCTCCTTCCATGAGGAAGCCGCCCTTTTGTGTGGAACATGC	1020
Qy	1021	TGCGTGTGGCAAAACATTTTGCAATGGAACAAGCTCTCACTAGGACGTGTTGTACATGA	1080
Db	1021	TGCGTGTGGCAAAACATTTTGCAATGGAACAAGCTCTCACTAGGACGTGTTGTACATGA	1080
Qy	1081	TTCGCAACAAGAAATGAAAGCTCAAACTCAAAAANTCTGTGAAAAACGAGATTTGGC	1140
Db	1081	TTCGCAACAAGAAATGAAAGCTCAAACTCAAAAANTCTGTGAAAAACGAGATTTGGC	1140
Qy	1141	CTCTCATCTCAGTGTATATATCCCTCCCAAAAGAAACAAAGGCAAGCTTATCTTTGTG	1200
Db	1141	CTCTCATCTCAGTGTATATATCCCTCCCAAAAGAAACAAAGGCAAGCTTATCTTTGTG	1200
Qy	1201	TCAAAACGAGAGACTACCCCAACTGTGTGGAAGACAAGATGCTCTGACAGTTGACACT	1260
Db	1201	TCAAAACGAGAGACTACCCCAACTGTGTGGAAGACAAGATGCTCTGACAGTTGACACT	1260
Qy	1261	TACCCCTGGCTAA 1273	
Db	1261	TACCCCTGGCTAA 1273	
RESULT 5			
US-09-644-868-9602			
; Sequence 9602, Application US/09644868			
; GENERAL INFORMATION:			
; APPLICANT: White, David			
; APPLICANT: Holtzman, Douglas A.			
; APPLICANT: Pan, Yang			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
; TITLE OF INVENTION: THEREFOR			
; FILE REFERENCE: 1600.1168-001			
; CURRENT APPLICATION NUMBER: US/09/644,868			
; CURRENT FILING DATE: 2000-08-28			
; PRIOR APPLICATION NUMBER: 60/151,063			
; PRIOR FILING DATE: 1999-08-27			
; NUMBER OF SEQ ID NOS: 10075			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 9602			
; LENGTH: 1597			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)..(1597)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-644-868-9602			
Query Match 99.7%; Score 1269.8; DB 25; Length 1597;			
Best Local Similarity 99.8%; Pred. No. 1.1e-276;			

	Matches	1271	Conservative	0	Mismatches	2	Indels	0	Gaps	0
QY	1	ATGCCAGACAGGGGGCCGACGCGGGGGCGTGCTCTGTGTACGCGCGCGCTCCCGAGT	60							
Db	18	ATGCGACAGACGGGGCCGACGCGGGGGCGTGCTGTGTATGACGCGCGCTCCGGAAT	77							
QY	61	GTCGCCGCTCCCGGGAAGGTTACAGAGGAGCCGTGGGCGGGCGCTGCCGCTCCGCG	120							
Db	78	GTCGCCGCTCCCGGGAAGGTTACAGAGGAGCCGTGGGCGGGCGCGCTGCCGCTCCGCG	137							
QY	121	ACGTCTCTGGACGCGGGGACGGCGCTGGGCGCTGGAGGCGCGGGCGCTCGA	180							
Db	138	ACGTCTCTGGACGCGGGGACGGCGCTGGGCGCTGGAGGCGCGGGCGCTCGA	197							
QY	181	TCCGCGCGCGCTGTGTGCGCCGAGTCCGTTGCTCTTGAACATCGCGAGCGCTTCAATTC	240							
Db	198	TCCGCGCGCGCTGTGTGCGCCGAGTCCGTTGCTCTTGAACATCGCGAGCGCTTCAATTC	257							
QY	241	ACCGCGCGAGAGCTCAGTCCGACCCCGCGCGCGCGCTTCCAGAGGTTCAATTC	300							
Db	258	ACCGCGCGAGAGCTCAGTCCGACCCCGCGCGCGCGCTTCCAGAGGTTCAATTC	317							
QY	301	CTTCCTTCCCTGACTCGAGGCGCAATTACAGCAAAAGCTTGAAAGCTTGACCGGCACCTGG	360							
Db	318	CTTCCTTCCCTGACTCGAGGCGCAATTACAGCAAAAGCTTGAAAGCTTGACCGGCACCTGG	377							
QY	361	CAAGACACAGGGGGGAGAGACATTGTGTGTGAAGGTGTGGCAAGGCTTCAT	420							
Db	378	CAAGACACAGGGGGGAGAGACATTGTGTGTGAAGGTGTGGCAAGGCTTCAT	437							
QY	421	CAGGAGCTACCATCTGAGCGCGCACATTTCTGACTCACAGAGAGAAAACCGTTGTTTG	480							
Db	438	CAGGAGCTACCATCTGAGCGCGCACATTTCTGACTCACAGAGAGAAAACCGTTGTTTG	497							
QY	481	TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATCAACCTTGGAAGAACATTTTGA	540							
Db	498	TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATCAACCTTGGAAGAACATTTTGA	557							
QY	541	ACGCAAACTGAAAAATCACAAAAACAATATATATGCACTTTTGAAGACTGTGAAGAAG	600							
Db	558	ACGCAAACTGAAAAATCACAAAAACAATATATATGCACTTTTGAAGACTGTGAAGAAG	617							
QY	601	CTTTAAGAAACATCAGACAGCTGAAAAATCCATAGTGCCAGCATACCAATGAACCTCTATT	660							
Db	618	CTTTAAGAAACATCAGACAGCTGAAAAATCCATAGTGCCAGCATACCAATGAACCTCTATT	677							
QY	661	CAAGTGTACCCAGGAAGATGTGGGAACACTTTCATCACCAGCAACTGAACGACA	720							
Db	678	CAAGTGTACCCAGGAAGATGTGGGAACACTTTCATCACCAGCAACTGAACGACA	737							
QY	721	TGCCAAGGCGCCAGAGGCGTATGTATGTCAAAAAGATGTCTCTTGTGGCAAAAACATG	780							
Db	738	TGCCAAGGCGCCAGAGGCGTATGTATGTCAAAAAGATGTCTCTTGTGGCAAAAACATG	797							
QY	781	GACGGAACCTTCTGAACAATGTAGAGAAACCCATAAAGAGAAATACTATGTGAAGTATG	840							
Db	798	GACGGAACCTTCTGAACAATGTAGAGAAACCCATAAAGAGAAATACTATGTGAAGTATG	857							
QY	841	CCGGAACAATTTTAAACGCAAAAGATTACCTTAAGCAACACATGAATAACTCATGCCACA	900							
Db	858	CCGGAACAATTTTAAACGCAAAAGATTACCTTAAGCAACACATGAATAACTCATGCCACA	917							
QY	901	AAGGGAATGTATTCGCTGTCCAAGAAAGCGTGGAAACACTTACTACTGTGTTTAA	960							
Db	918	AAGGGAATGTATTCGCTGTCCAAGAAAGCGTGGAAACACTTACTACTGTGTTTAA	977							
QY	961	TCTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTTGTGTGAACATATC	1020							
Db	978	TCTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTTGTGTGAACATATC	1037							
QY	1021	TGGCTGTGGCAAAACATTTTGCATGAACAAGTCTCACTAGGCATGTGTGTATATGA	1080							
Db	1038	TGGCTGTGGCAAAACATTTTGCATGAACAAGTCTCACTAGGCATGTGTGTATATGA	1097							

OY	1081	TCCTACCAAGAAAGAAATGAAACCTCAAAAGTCAAAAATCTGTGAAAAAAGGAGTTGGC	1140
Db	1098	TCCTACCAAGAAAGAAATGAAACCTCAAAAGTCAAAAATCTGTGAAAAAAGGAGTTGGC	1157
OY	1141	CTCTCATCTCAAGTGAATATATCTCCCTCCCAAAAGGAAGGCAAGCTTATCTTTGTG	1200
Db	1158	CTCTCATCTCAAGTGAATATATCTCCCTCCCAAAAGGAAGGCAAGGCTTATCTTTGTG	1217
OY	1201	TCAAAAGGAGAGTACCCCACTGTGTGGAGAGACAAGATGCTCTGACAGTTGGACGTACT	1260
Db	1218	TCAAAAGGAGAGTACCCCACTGTGTGGAGAGACAAGATGCTCTGACAGTTGGACGTACT	1277
OY	1261	TACCTTGGCTAA 1273	
Db	1278	TACCTTGGCTAA 1290	
RESULT 6			
US-09-652-126-8862			
: Sequence 8862, Application US/09652126			
: GENERAL INFORMATION:			
: APPLICANT: Shyjan, Andrew W.			
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
: FILE OF INVENTION: THEREFOR			
: FILE REFERENCE: 1600.1185-001			
: CURRENT APPLICATION NUMBER: US/09/652,126			
: PRIOR FILING DATE: 2000-08-30			
: PRIOR APPLICATION NUMBER: 60/151,132			
: PRIOR FILING DATE: 1999-08-30			
: NUMBER OF SEQ ID NOS: 10051			
: SOFTWARE: FASTSEQ for Windows Version 4.0			
: SEQ ID NO 8862			
: LENGTH: 1597			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (1)..(1597)			
: OTHER INFORMATION: n = A,T,C or G			
US-09-652-126-8862			
Query Match 99.7%; Score 1269.8; DB 25; Length 1597;			
Best Local Similarity 99.8%; Pred. No. 1.1e-278;			
Matches 1271; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
OY	1	ATGGCCACACAGCGGGCGCCGACCGCGGGCGCGTCCCTGTGTACCCGCGCGCTCCCGAAGT	60
Db	18	ATGGCCACACAGCGGGCGCCGACCGCGGGCGGTGTCCTGTGTACCCGCGCGCTCCCGAAGT	77
OY	61	GTGCGCGGCGTCCGCGGAAGTTTACACAGAGACCGCTGGCGCGGGCGCGCGGTTCCGGC	120
Db	78	GTGCGCGGCGTCCGCGGAAGTTTACACAGAGAGCCGTGGCGCGGGCGCGCGGTTCCGGC	137
OY	121	ACGTGTCTTCGCGACGTGGCAGCGCGCCGTGGCTTGGAGAGCGCGCGCCCTGGA	180
Db	138	ACGTGTCTTCGCGACGTGGCAGCGCGCCGTGGCTTGGAGAGCGCGCGCCCTGGA	197
OY	181	TCCCGCGCGCGTGTGTCCCGGAGTCCGTGTCTTGTGACATCGCCGACGCTTATTGC	240
Db	198	TCCCGCGCGCGTGTGTCCCGGAGTCCGTGTCTTGTGACCAATCGCCGACGCTTATTGC	257
OY	241	AGCGCGGAGAGACTAGCTCCGACCGCGCGCGCGCGCTTCCAGAGGTTGATCG	300
Db	258	AGCGCGGAGAGACTAGCTCCGACCGCGCGCGCGCGCTTCCAGAGGTTGATCTG	317
OY	301	CTCTTCCCTGACTGACGCGCCCAATTACAGCAAAAGCTTGAAGCTTGAACGCGCACCTGTG	360
Db	318	CTCTTCCCTGACTGACGCGCCCAATTACAGCAAAAGCTTGAAGGCTTGAACGCGCACCTGTG	377
OY	361	CAAGCAACAGGGGGAGAGACCATTTTCTTTGTGTACTATGAAGGTTGTGCAAGGCTTCAT	420
Db	378	CAAGCAACAGGGGGAGAGACCATTTTGTGTGTGTACTATGAAGGTTGTGCAAGGCTTCAT	437

OY	421	CAGBACATACCATTGTGAGCCGCCAACATTTGACTGCACAGGAGAAAGCCGTTGGTG	480
OY	421	CAGBACATACCATTGTGAGCCGCCAACATTTGACTGCACAGGAGAAAGCCGTTGGTG	480
Db	438	CAGGACATACCATTGTGAGCCGCCACATTTGTGACTCACAGAGAAAGCCGTTGGTG	497
OY	481	TGCAGCCACTGGCTGTATCATAAAATTTCAACACAAANTCAAACTTGAAGAAACATTTTGA	540
Db	481	TGCAGCCACTGGCTGTATCATAAAATTTCAACACAAANTCAAACTTGAAGAAACATTTTGA	540
Db	498	TGCAGCCAAATGGCTGTATCAAAAATTTCAACACAAANTCAAACTTGAAGAAACATTTTGA	557
OY	541	ACGCAAAACATGAAAATCATCAACAAAACAAATATATATGACAGTTTGAAGACTGTAGAAGAC	600
Db	541	ACGCAAAACATGAAAATCATCAACAAAACAAATATATATGACAGTTTGAAGACTGTAGAAGAC	600
Db	558	ACGCAAAACATGAAAATCAACAAAACAAATATATATGACAGTTTGAAGACTGTAGAAGAC	617
OY	601	CTTTAAGAAACATGAGAGAGCTGAAAATCCATCATGTCGCCAGCATACCCAATGAACCTCTATT	660
Db	601	CTTTAAGAAACATGAGAGAGCTGAAAATCCATCATGTCGCCAGCATACCCAATGAACCTCTATT	660
Db	618	CTTTAAGAAACATGAGAGAGCTGAAAATCCATCATGTCGCCAGCATACCCAATGAACCTCTATT	677
OY	661	CAAGTGTACCCAGAGAGATGTGGGAAACACTTGGCATCACCCAGCAGCTGAAGCAGCA	720
Db	678	CAAGTGTACCCAGAGAGATGTGGGAAACACTTGGCATCACCCAGCAGCTGAAGCAGCA	737
OY	721	TGCCAAGGCCACGAGGCGTATGTATGTCAAAAAGATGTTCCCTTTGGCAAAAACATG	780
Db	738	TGCCAAGGCCACGAGGCGTATGTATGTCAAAAAGATGTTCCCTTTGGCAAAAACATG	797
OY	781	GACGGAACCTTGTGAACATGTGAGAGAAACCCATAAGAGGAAATACATGTGTGAAGTATG	840
Db	798	GACGGAACCTTGTGAACATGTGAGAGAAACCCATAAGAGGAAATACATGTGTGAAGTATG	857
OY	841	CCGGAACCATTTAAACGCAAGAATTACCTTAAGCAACACATGAAAACACTGACCCGAGA	900
Db	858	CCGGAACCATTTAAACGCAAGAATTACCTTAAGCAACACATGAAAACACTGACCCGAGA	917
OY	901	AAGGATGTATGTGCGTCGTCCAGAGAGGCGTGGAAAGCCTATACTACTGTGTTAA	960
Db	918	AAGGATGTATGTGCGTCGTCCAGAGAGGCGTGGAAAGCCTATACTACTGTGTTAA	977
OY	961	TCTCCAAAAGCCATPTCCTCTCCTTCCATGAGGAAACGCCCTTTTGTGTGGAACATGC	1020
Db	978	TCTCCAAAAGCCATPTCCTCTCCTTCCATGAGGAAACGCCCTTTTGTGTGGAACATGC	1037
OY	1021	TGCGTGTGGCAAAACATTTGCAATGAANCAAGCTCTCAGGACATGCTGTTGTACATGA	1080
Db	1038	TGCGTGTGGCAAAACATTTGCAATGAANCAAGCTCTCAGGACATGCTGTTGTACATGA	1097
OY	1081	TCCGACAAAGAAAGAAATGAAAGCTCAAAAGTCAAAATCTCGGAAAAACGGAGTTTGGC	1140
Db	1098	TCCGACAAAGAAAGAAATGAAAGCTCAAAAGTCAAAATCTCGGAAAAACGGAGTTTGGC	1157
OY	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGAAACGAGCAAGGCTTATCTTTTGTG	1200
Db	1158	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGAAACGAGCAAGGCTTATCTTTTGTG	1217
OY	1201	TCCAAAACGAGAGATCACCCCAACTGTGTGAGACAAAGATGCTCTGCACAGTTGCAGTACT	1260
Db	1218	TCCAAAACGAGAGATCACCCCAACTGTGTGAGACAAAGATGCTCTGCACAGTTGCAGTACT	1277
OY	1261	TACCCCTGGCTAA 1273	
Db	1278	TACCCCTGGCTAA 1290	
RESULT 7			
US-09-652-814-10703			
; Sequence 10703, Application US/09652814			
; GENERAL INFORMATION:			
; APPLICANT: Holtzman, Douglas A.			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
; FILE REFERENCE: 1600.1191-001			
; CURRENT APPLICATION NUMBER: US/09/652,814			
; PRIORITY FILING DATE: 2000-08-31			
; PRIORITY APPLICATION NUMBER: 60/152,109			

```
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 10797
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10703
; LENGTH: 1597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1597)
; OTHER INFORMATION: n = A,T,C or G
US-09-652-814-10703
```

```
Query Match          99.7%; Score 1269.8; DB 25; Length 1597;
Best Local Similarity 99.8%; Pred. No. 1.1e-278;
Matches 1271; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 ATGGCGACAGCGGCGGCGGCGGCGGCTGTGATGACCGCGCGCTCCCGAAGT 60
    |||||||
DB 18 ATGGCGACAGCGGCGGCGGCGGCGGCTGTGATGACCGCGCGCTCCCGAAGT 77
    |||||||
OY 61 GTGCGGCGCTGCGCGGAGGTTTCAGAGGAGCGGCGGCGGCGGCGGCTCCGCGC 120
    |||||||
DB 78 GTGCGGCGCTGCGCGGAGGTTTCAGAGGAGCGGCGGCGGCGGCGGCTCCGCGC 137
    |||||||
OY 121 ACGGTCTCGGACGTGGACGCGCGGCGGCGGCTGGGCTTGGAGGCGCGGCGCTGGA 180
    |||||||
DB 138 ACGGTCTCGGACGTGGACGCGCGGCGGCGGCTGGGCTTGGAGGCGCGGCGCTGGA 197
    |||||||
OY 181 TCGCGGCGGCGGTGTGCGCGGAGTGTGTGCTTACCATCGCGACGCGCTTCAATGC 240
    |||||||
DB 198 TCGCGGCGGCGGTGTGCGCGGAGTGTGTGCTTACCATCGCGACGCGCTTCAATGC 257
    |||||||
OY 241 AGCGGCGGAGAGTCACTCCGACCGCGCGCGCGCGGCTTCCAGGAGTTCATCTG 300
    |||||||
DB 258 AGCGGCGGAGAGTCACTCCGACCGCGCGCGCGCGGCTTCCAGGAGTTCATCTG 317
    |||||||
OY 301 CTCCTTCCCTGACTGACGCGGCAATTACAGCAAAAGCTTGAAGCTTACGCGCACTGTG 360
    |||||||
DB 318 CTCCTTCCCTGACTGACGCGGCAATTACAGCAAAAGCTTGAAGCTTACGCGCACTGTG 377
    |||||||
OY 361 CAAGACACAGGCGGAGAGACCATTTGTTGTGACTATGAAAGGTGTGGCAAGCGCTTCAT 420
    |||||||
DB 378 CAAGACACAGGCGGAGAGACCATTTGTTGTGACTATGAAAGGTGTGGCAAGCGCTTCAT 437
    |||||||
OY 421 CAGGAGTACCATGTGAGCGCGCACATTTGACTCACACAGGAGAAAAGCGCTTGTGTTG 480
    |||||||
DB 438 CAGGAGTACCATGTGAGCGCGCACATTTGACTCACACAGGAGAAAAGCGCTTGTGTTG 497
    |||||||
OY 481 TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATTCAAAATTCAGAAACATTTTGA 540
    |||||||
DB 498 TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATTCAAAATTCAGAAACATTTTGA 557
    |||||||
OY 541 ACGCAAAATGAAAAATCAACAAAATATATATATGCAAGTTTGAAGACTGTAAAGAAC 600
    |||||||
DB 558 ACGCAAAATGAAAAATCAACAAAATATATATATGCAAGTTTGAAGACTGTAAAGAAC 617
    |||||||
OY 601 CTTTAAGAAACATCAGAGCTGAAATTCATCATGSCAGATACCAATGAACCTCTATT 660
    |||||||
DB 618 CTTTAAGAAACATCAGAGCTGAAATTCATCATGSCAGATACCAATGAACCTCTATT 677
    |||||||
OY 661 CAATGTACCCAGAGAGATGTGGAAAACACTTTCATCACCCAGCAAGCTGAAACGACA 720
    |||||||
DB 678 CAATGTACCCAGAGAGATGTGGAAAACACTTTCATCACCCAGCAAGCTGAAACGACA 737
    |||||||
OY 721 TGCAGAGCCGACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 780
    |||||||
DB 738 TGCAGAGCCGACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 797
    |||||||
OY 781 GAGGAACTTGTGAACAATGAGAGAAACCATTAAGAGAAATACATGTGAAGATG 840
    |||||||
DB 798 GAGGAACTTGTGAACAATGAGAGAAACCATTAAGAGAAATACATGTGAAGATG 857
    |||||||
```

```
OY 841 CCGGAAAACATTTAAACGCAAAAGATTACCTTAAGCAACACATGAAAAATCATGCCGACA 900
    |||||||
DB 858 CCGGAAAACATTTAAACGCAAAAGATTACCTTAAGCAACACATGAAAAATCATGCCGACA 917
    |||||||
OY 901 AAGGATGTATGTGCGCTGTCCAGAGAGGCTGTGGAAGAACTATATCTGTGTTAA 960
    |||||||
DB 918 AAGGATGTATGTGCGCTGTCCAGAGAGGCTGTGGAAGAACTATATCAACTGTGTTAA 977
    |||||||
OY 961 TCTCCAAAGCCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
    |||||||
DB 978 TCTCCAAAGCCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1037
    |||||||
OY 1021 TGGCTGTGCAAAAACATTTGCAATGAACAAGCTCTCAGTACGATGCTTGTATGTA 1080
    |||||||
DB 1038 TGGCTGTGCAAAAACATTTGCAATGAACAAGCTCTCAGTACGATGCTTGTATGTA 1097
    |||||||
OY 1081 TCCGTGACAGAGAAATGAAAGCTCAAGTCAAAAAATCTCTGTGAAAAACGAGTTGGC 1140
    |||||||
DB 1098 TCCGTGACAGAGAAATGAAAGCTCAAGTCAAAAAATCTCTGTGAAAAACGAGTTGGC 1157
    |||||||
OY 1141 CTCCTCATCTCAGTGAATATATCCCTCCCAAAAGGAAACAGGCGCAAGCTTATCTTGTG 1200
    |||||||
DB 1158 CTCCTCATCTCAGTGAATATATCCCTCCCAAAAGGAAACAGGCGCAAGCTTATCTTGTG 1217
    |||||||
OY 1201 TCAAAAGGAGAGTCAACCAACTGTGTGAAGAACAAAGATCTCTGACAGTTGCAGTACT 1260
    |||||||
DB 1218 TCAAAAGGAGAGTCAACCAACTGTGTGAAGAACAAAGATCTCTGACAGTTGCAGTACT 1277
    |||||||
OY 1261 TACCCTTGGCTAA 1273
    |||||||
DB 1278 TACCCTTGGCTAA 1290
    |||||||
```

RESULT 8

```
US-09-716-473-1934
; Sequence 1934, Application US/09716473
; GENERAL INFORMATION:
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.2043-001
; CURRENT APPLICATION NUMBER: US/09/716.473
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/166,502
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 2933
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1597)
; OTHER INFORMATION: n = A,T,C or G
US-09-716-473-1934
```

```
Query Match          99.7%; Score 1269.8; DB 28; Length 1597;
Best Local Similarity 99.8%; Pred. No. 1.1e-278;
Matches 1271; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 ATGGCGACAGCGGCGGCGGCGGCGGCTGTGATGACCGCGCGCTCCCGAAGT 60
    |||||||
DB 18 ATGGCGACAGCGGCGGCGGCGGCGGCGGCTGTGATGACCGCGCGCTCCCGAAGT 77
    |||||||
OY 61 GTGCGGCGCTGCGCGGAGGTTTCAGAGGAGCGGCGGCGGCGGCGGCTCCGCGC 120
    |||||||
DB 78 GTGCGGCGCTGCGCGGAGGTTTCAGAGGAGCGGCGGCGGCGGCGGCTCCGCGC 137
    |||||||
OY 121 ACGGTCTCGGACGTGGACGCGCGGCGGCTGGGCTTGGAGGCGCGGCGCTTGA 180
    |||||||
DB 138 ACGGTCTCGGACGTGGACGCGCGGCGGCTGGGCTTGGAGGCGCGGCGCTTGA 197
    |||||||
```

QY	181	TCGCCCGCCGTGGTGTCCGCGAATGTCGATGTCTGCTTGACACATTCGCCGACGCTTATTTGC	240
Dp	198	TCGCCCGCCGTGGTGTCCGCGAATGTCGATGTCTGCTTGACACATTCGCCGACGCTTATTTGC	257
QY	241	AGCCGGGAGAGACTACACTCCGACCCCGCGCCGCCGCGTTCGACGAGATTTATCTG	300
Dp	258	AGCCGGGAGAGACTACACTCCGACCCCGCGCCGCCGCGTTCGACGAGATTTATCTG	317
QY	301	CTTCCTTCCCTBACTGCAAGCCCAATTAACGAAGCTTGGAAGCTTGACGCGCACCTGTG	360
Dp	318	CTTCCTTCCCTBACTGCAAGCCCAATTAACGAAGCTTGGAAGCTTGACGCGCACCTGTG	377
QY	361	CAACGACACGGGGAGAGACATTTGTTGTGCTATGAAGGGGTGGCAACGCTTCAT	420
Dp	378	CAACGACACGGGGAGAGACATTTGTTGTGCTATGAAGGGGTGGCAACGCTTCAT	437
QY	421	CAGGACTACCATCTGAGCCGCCACATTTCTGACTCACAGAGAGAAAACCGTTGTTTG	480
Dp	438	CAGGACTACCATCTGAGCCGCCACATTTCTGACTCACAGAGAGAAAACCGTTGTTTG	497
QY	481	TGCGAGCCACTGGGCTGTATCAAAAATTCAAAACAAAATCAAACTTGAAGAAACATTTTGA	540
Dp	498	TGCGAGCCACTGGGCTGTATCAAAAATTCAAAACAAAATCAAACTTGAAGAAACATTTTGA	557
QY	541	ACCGAAACATGAAAAATACAAAAAACAATATATATCAGTTTGAAGACTGTAGAGAAAC	600
Dp	558	ACCGAAACATGAAAAATACAAAAAACAATATATATCAGTTTGAAGACTGTAGAGAAAC	617
QY	601	CTTTTAAGAAACATATACGACGTGAAATCCATCAGTCCGACATACCAATGAACCTCTATT	660
Dp	618	CTTTTAAGAAACATATACGACGTGAAATCCATCAGTCCGACATACCAATGAACCTCTATT	677
QY	661	CAAGTGTACCCAGGAGAGAGATGTGGGAAACACTTGTGATCACCCAGCAACCTGAACGACA	720
Dp	678	CAAGTGTACCCAGGAGAGAGATGTGGGAAACACTTGTGATCACCCAGCAACCTGAACGACA	737
QY	721	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCCCTTTGTGGCAAAAACATG	780
Dp	738	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCCCTTTGTGGCAAAAACATG	797
QY	781	GACGGAACCTTCTGAAACATGTGAGAGAAACCCCTAAGAGGAAATACATATGTGAAGTATG	840
Dp	798	GACGGAACCTTCTGAAACATGTGAGAGAAACCCCTAAGAGGAAATACATATGTGAAGTATG	857
QY	841	CCGGAACCATTTAAAGGCAAGAATTAACCTTAAGCAACACATGAAGAACTCATGCCCCAGA	900
Dp	858	CCGGAACCATTTAAAGGCAAGAATTAACCTTAAGCAACACATGAAGAACTCATGCCCCAGA	917
QY	901	AAGGATGTATGTGCGGTGTCCAAAGAGCGTGTGGAACAACTTACTACTGTGTTAA	960
Dp	918	AAGGATGTATGTGCGGTGTCCAAAGAGCGTGTGGAACAACTTACTACTGTGTTAA	977
QY	961	TCTCCAAAGCCATTCCTCTCTCTTCATGAGAAACCCGCTTTTGTGTGGAACATGC	1020
Dp	978	TCTCCAAAGCCATTCCTCTCTCTTCATGAGAAACCCGCTTTTGTGTGGAACATGC	1037
QY	1021	TGCGTGTGGCAAAACATTTGCAATGAACAAAGCTCTCACTAGGCATGCTGTTGATCATCA	1080
Dp	1038	TGCGTGTGGCAAAACATTTGCAATGAACAAAGCTCTCACTAGGCATGCTGTTGATCATCA	1097
QY	1081	TCTCGACAAAGAAATTAAGAAAGCTCAAAAGTCAAAATAATCTCGTGAAGAAAACGAGTTTGGC	1140
Dp	1098	TCTCGACAAAGAAATTAAGAAAGCTCAAAAGTCAAAATAATCTCGTGAAGAAAACGAGTTTGGC	1157
QY	1141	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGGAACAAAGGCAAGGCTTATCTTTGTG	1200
Dp	1158	CTCTCATCTCAGTGTGATATATCCCTCCCAAAAGGAACAAAGGCAAGGCTTATCTTTGTG	1217
QY	1201	TCAAAACGGAGAGTCAACCCCACTGTGTGGAAGCAAGATGCTCTGTGACAGTTTGACATCT	1260
Dp	1218	TCAAAACGGAGAGTCAACCCCACTGTGTGGAAGCAAGATGCTCTGTGACAGTTTGACATCT	1277

QY	1261	TACCGTTGGCTAA	1273
DB	1278	TACCGTTGGCTAA	1290
RESULT 9			
US-09-716-990-1322			
; Sequence 1322, Application US/09716990			
; GENERAL INFORMATION:			
; APPLICANT: Gearing, David P.			
; APPLICANT: Holtzman, Douglas A.			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
; TITLE OF INVENTION: THEREFOR			
; FILE REFERENCE: 1600.2039-001			
; CURRENT APPLICATION NUMBER: US/09/716,990			
; CURRENT FILING DATE: 2000-11-21			
; PRIOR APPLICATION NUMBER: US 60/166,926			
; PRIOR FILING DATE: 1999-11-22			
; NUMBER OF SEQ ID NOS: 1691			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 1322			
; LENGTH: 1597			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc-feature			
; LOCATION: (1)...(1597)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-716-990-1322			
Query Match 99.7%; Score 1269.8; DB 28; Length 1597;			
Best Local Similarity 99.8%; Pred.No.1,1e-278;			
Matches 1771; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	ATGCGCAGAGCGGGCGCCGACCGCGGGCGGCTGCTGTGACCGCGCGGCTCCGGAAGT	60
DB	18	ATGCGCAGAGCGGGCGCCGACCGCGGGCGGCTGCTGTGACCGCGCGGCTCCGGAAGT	77
QY	61	GTGCCGGCTCGCGGAAAGTTACAGAGGAGCCGTGGGCGGGCGCGCGGTTCCCGGC	120
DB	78	GTGCCGGCTCGCGGAAAGTTACAGAGGAGCCGTGGGCGGGCGCGCGGTTCCCGGC	137
QY	121	ACGTGTTCTGGCAGCTGGCAGCGCGGCTGGCCCTGGGCTTGGAGCGCCGCGCCCTTGA	180
DB	138	ACGTGTTCTGGCAGCTGGCAGCGCGGCTGGCCCTGGGCTTGGAGCGCGCCCTTGA	197
QY	181	TCCGCGGGCGCTGTGTCGCCGAGTCGGGTGTCGCTTGACATCGCGGAGCGGTTCAATTC	240
DB	198	TCCGCGGGCGCTGTGTCGCCGAGTCGGGTGTCGCTTGACATCGCGGAGCGGTTCAATTC	257
QY	241	ACCGCGGCGAGAGCTCAGCTCCGACCCCGCGCGCCCGGCTTCCAGAGGTTGATCTG	300
DB	258	ACCGCGGCGAGAGCTCAGCTCCGACCCCGCGCGCCCGGCTTCCAGAGGTTGATCTG	317
QY	301	CTTCCTTCCCTGACTGCAGCGGCCAATTACAGCAAGCCTTGAAGCTTGAACGCGCACTGTG	360
DB	318	CTTCCTTCCCTGACTGCAGCGGCCAATTACAGCAAGCCTTGAAGCTTGAACGCGCACTGTG	377
QY	361	CAAGACAGAGGGGGAGAGACATTTGTTGTGACTATGAAGGCTGGCAAGGCCCTCAT	420
DB	378	CAAGACAGAGGGGGAGAGACATTTGTTGTGACTATGAAGGCTGGCAAGGCCCTCAT	437
QY	421	CAGGAGCTACATCTGAGAGCGCCACATTTGACTACACAGAGAGAAAAGCCCTTTGTTTG	480
DB	438	CAGGAGCTACATCTGAGAGCGCCACATTTGACTACACAGAGAGAAAAGCCCTTTGTTTG	497
QY	481	TGCAGCACTGGCTGTGATCAAAAATTCAACACAAAATCAAACTTTGAAGAAACATTTTGA	540
DB	498	TGCAGCAATGCGTGTGATCAAAAATTCAACACAAAATCAAACTTTGAAGAAACATTTTGA	557
QY	541	ACGCAAACTGAAAATCAACAAAACAAATATATATGCACTTTTGAAGACTGTAAAGAAC	600
DB	558	ACGCAAACTGAAAATCAACAAAACAAATATATATGCACTTTTGAAGACTGTAAAGAAC	617

```
QY 601 CTTTAAAGAACATCAGCAGCTGAATAATCATCATGCGCAGCATACCAATGAACTTATT 660
    |||||||
Db 618 CTTTAAAGAACATCAGCAGCTGAATAATCATCATGCGCAGCATACCAATGAACTTATT 677
QY 661 CAACTGTACCCAGAGAGAGTGGGAAACACTTTGCATCACCCAGCAAGCTGAAGACA 720
    |||||||
Db 678 CAACTGTACCCAGAGAGATGGGAAACACTTTGCATCACCCAGCAAGCTGAAGACA 737
QY 721 TGGCAAGGCCAGCAGGCTGTATGTATCAAAAAGATGTTCCCTTTGGCAAAAACATG 780
    |||||||
Db 738 TGGCAAGGCCAGCAGGCTGTATGTATGTCAAAAAGATGTTCCCTTTGGCAAAAACATG 797
QY 781 GACGGAATCTCTGAACATGTGAGAGAAACCATAAGAGAAATACTATGTGAAGTATG 840
    |||||||
Db 798 GACGGAATCTCTGAACATGTGAGAGAAACCATAAGAGAAATACTATGTGAAGTATG 857
QY 841 CCGGAAAAATTTAAAGCAAGATTAACCTTAAGCAACACATGAAACATGCCCCAGA 900
    |||||||
Db 858 CCGGAAAAATTTAAAGCAAGATTAACCTTAAGCAACACATGAAACATGCCCCAGA 917
QY 901 AAGGATGTATGTGCGCTGTCCAGAGAGGCTGTGGAAGAACTATACTGTGTTTAA 960
    |||||||
Db 918 AAGGATGTATGTGCGCTGTCCAGAGAGGCTGTGGAAGAACTATACTGTGTTTAA 977
QY 961 TCTCAAGGCCATATCCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020
    |||||||
Db 978 TCTCAAGGCCATATCCTCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1037
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCACTAGGCATGCTGTTGTACATGA 1080
    |||||||
Db 1038 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCACTAGGCATGCTGTTGTACATGA 1097
QY 1081 TCCGACAGAGAGAAATGAAGCTCAAAAGTCAAAAATCTGTGAAAAACGAGTTTGGC 1140
    |||||||
Db 1098 TCCGACAGAGAGAAATGAAGCTCAAAAGTCAAAAATCTGTGAAAAACGAGTTTGGC 1157
QY 1141 CTCATCATCTAGTGGATATATCCCTCCCAAAAGAGCAAGGCAAGGCTTATCTTTGTG 1200
    |||||||
Db 1158 CTCATCATCTAGTGGATATATCCCTCCCAAAAGAGCAAGGCAAGGCTTATCTTTGTG 1217
QY 1201 TCAAAAGAGAGATGACCCCAACTGTGTGAAGACAGAGATGCTGACAGTTGACATCT 1260
    |||||||
Db 1218 TCAAAAGAGAGATGACCCCAACTGTGTGAAGACAGAGATGCTGACAGTTGACATCT 1277
QY 1261 TACGCTGGCTAA 1273
Db 1278 TACGCTGGCTAA 1290
```

```
RESULT 10
US-09-721-589-6336
: Sequence 6336, Application US/09721589
: GENERAL INFORMATION:
: APPLICANT: Gearing, David P.
: APPLICANT: Villalva, Jean-Luc
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600,2045-001
: CURRENT APPLICATION NUMBER: US/09/721,589
: PRIOR FILING DATE: 2000-11-22
: PRIOR APPLICATION NUMBER: 60/167,380
: NUMBER OF SEQ ID NOS: 1999-11-24
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6336
: LENGTH: 1597
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(1597)
: OTHER INFORMATION: n = A,T,C or G
```

```
US-09-721-589-6336
Query Match          99.7%: Score 1269.8: DB 29: Length 1597:
Best Local Similarity 99.8%: Pred. No. 1,1e-278:
Matches 1271: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

QY 1 ATGCGCAGCAGCGGCGCCGACAGCGGCGGCGGTGCTGTGATGACCCGCGCGCTCCCGAAGT 60
    |||||||
Db 18 ATGCGCAGCAGCGGCGGCGGCGGCGGCGGCGGTGCTGTGATGACCCGCGCGCTCCCGAAGT 77
QY 61 GTGCGGCGGTGCGCGGAGAGTTACAGCAGGAGCCGTGCGCGCGCGCGGCTTCCCGC 120
    |||||||
Db 78 GTGCGGCGGTGCGCGGAGAGTTACAGCAGGAGCCGTGCGCGCGCGCGGCTTCCCGC 137
QY 121 ACGTGTCTGCGCAGCTGCGAGCGCGCGCTGCGCTGCGCGCTGCGCGCGCGCGCGCGCG 180
    |||||||
Db 138 ACGTGTCTGCGCAGCTGCGAGCGCGCGCTGCGCTGCGCGCTGCGCGCGCGCGCGCGCG 197
QY 181 TCGCGCGCGCGGTGCGCGAGTGTGCTGTGCTGTGATGACCATGCGCGAGCGCTTCAATTGC 240
    |||||||
Db 198 TCGCGCGCGCGGTGCGCGAGTGTGCTGTGCTGTGATGACCATGCGCGAGCGCTTCAATTGC 257
QY 241 AGCGGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCGCTTCCAGAGGTTCAATCTG 300
    |||||||
Db 258 AGCGGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCGCTTCCAGAGGTTCAATCTG 317
QY 301 CTCCTTCCCTGTGCTGCGAGCGCGCAATTACAGCAAGCTTGAAGCTTACCGCGCATCTG 360
    |||||||
Db 318 CTCCTTCCCTGTGCTGCGAGCGCGCAATTACAGCAAGCTTGAAGCTTACCGCGCATCTG 377
QY 361 CAAGCAGCAGGCGGAGAGACCATTTGTTGTGACTATGAAGGGGTGGCAAGGCTTCAAT 420
    |||||||
Db 378 CAAGCAGCAGGCGGAGAGACCATTTGTTGTGACTATGAAGGGGTGGCAAGGCTTCAAT 437
QY 421 CAGGACTACCATCTGAGCGCGCACATTTGACTCACAGAGAGAAAGCCGTTTGTG 480
    |||||||
Db 438 CAGGACTACCATCTGAGCGCGCACATTTGACTCACAGAGAGAAAGCCGTTTGTG 497
QY 481 TGCAGCCACTGTGCTGTATCAAAAATTCACACAAATTCAAACTTGAAGAAACCTTTTGA 540
    |||||||
Db 498 TGCAGCCACTGTGCTGTATCAAAAATTCACACAAATTCAAACTTGAAGAAACCTTTTGA 557
QY 541 ACGCAACATGAAATATCAACAAACATATATATGCAAGTTTGAAGAGTGAAGAGAC 600
    |||||||
Db 558 ACGCAACATGAAATATCAACAAACATATATATGCAAGTTTGAAGAGTGAAGAGAC 617
QY 601 CTTTAAAGAACATCAGCAGCTGAATAATCATCATGCGCAGCATACCAATGAACTTATT 660
    |||||||
Db 618 CTTTAAAGAACATCAGCAGCTGAATAATCATCATGCGCAGCATACCAATGAACTTATT 677
QY 661 CAACTGTACCCAGAGAGTGTGGAAACCTTTGCATCACCCAGCAAGCTGAAGACA 720
    |||||||
Db 678 CAACTGTACCCAGAGAGTGTGGAAACCTTTGCATCACCCAGCAAGCTGAAGACA 737
QY 721 TGGCAAGGCCAGCAGGCTGTATGTATCAAAAAGATGTTCCCTTTGGCAAAAACATG 780
    |||||||
Db 738 TGGCAAGGCCAGCAGGCTGTATGTATGTCAAAAAGATGTTCCCTTTGGCAAAAACATG 797
QY 781 GACGGAATCTCTGAACATGTGAGAGAAACCATAAGAGAAATACTATGTGAAGTATG 840
    |||||||
Db 798 GACGGAATCTCTGAACATGTGAGAGAAACCATAAGAGAAATACTATGTGAAGTATG 857
QY 841 CCGGAAAAATTTAAAGCAAGATTAACCTTAAGCAACACATGAAACATGCCCCAGA 900
    |||||||
Db 858 CCGGAAAAATTTAAAGCAAGATTAACCTTAAGCAACACATGAAACATGCCCCAGA 917
QY 901 AAGGATGTATGTGCGCTGTCCAGAGAGGCTGTGGAAGAACTATACTGTGTTTAA 960
    |||||||
Db 918 AAGGATGTATGTGCGCTGTCCAGAGAGGCTGTGGAAGAACTATACTGTGTTTAA 977
QY 961 TCTCAAGGCCATATCCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020
    |||||||
Db 978 TCTCAAGGCCATATCCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1037
```


QY	1021	TGGCGTGGGCAAAACATTTCGATGGAACAAAGCTCTCACTAGGCATGCTGTTGTACATGA	1080
Db	1038	TGGCTGTGGCAAAACATTTCCAAATGAAACAAAGCTCTCACTAGGCATGCTGTTGTACATGA	1097
QY	1081	TCTGTACAGAGAGAAATGAAAGCTCAAAAGTCAAAAAATCTCGTGA AAAACGAGATTTGGC	1140
Db	1098	TCTGTACAGAGAGAAATGAAAGCTCAAAAGTCAAAAAATCTCGTGA AAAACGAGATTTGGC	1157
QY	1141	CTCTCATCTCAAGTGGATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTATCTTTGTG	1200
Db	1158	CTCTCATCTCAAGTGGATATATCCCTCCCAAAAGAAACAAAGGCAAGGCTATCTTTGTG	1217
QY	1201	TCAAAAGGAGAGTCAACCACATGTGTGGAGAACAAGATGCTCTGACAGTTGCAGTACT	1260
Db	1218	TCAAAAGGAGAGTCAACCACATGTGTGGAGAACAAGATGCTCTGACAGTTGCAGTACT	1277
QY	1261	TACCCCTTGGCTAA 1273	
Db	1278	TACCCCTTGGCTAA 1290	

```

RESULT 11
US-09-726-788-6896
? Sequence 6896, Application US/09726788
? GENERAL INFORMATION:
? APPLICANT: Gearing, David P.
? APPLICANT: Kingsbury, Gillian A.
? TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
? TITLE OF INVENTION: THEREFOR
? FILE REFERENCE: 1600.2051-001
? CURRENT APPLICATION NUMBER: US/09/726,788
? CURRENT FILING DATE: 2000-11-30
? PRIOR APPLICATION NUMBER: 60/168,131
? PRIOR FILING DATE: 1999-11-30
? NUMBER OF SEQ ID NOS: 7651
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 6896
? LENGTH: 1597
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)..(1597)
? OTHER INFORMATION: n = A,T,C OR G
US-09-726-788-6896

```

Query Match	99.7%	Score 1265.8	DB 29	Length 1597
Best Local Similarity	99.8%	Pred. No. 1.1e-278		
Matches 171	Conservative	0	Mismatches 2	Indels 0
			Gaps	0

Qy	1	ATGGCAGACACACGGCGCCCGAGACGGCGGGGCGTGTCTGTGTACCGCGCCGCTCCCGCAAGT	60
Db	18	ATGGCCACACACGGCGCCCGAGACGGCGGGGCGTGTGTGTACCGCGCCGCTCCCGCAAGT	77
Qy	61	GTCGCGGCGTGCAGAGTTTACAGACAGACGCGTGTGGCGCGGGCGGGCGGTTCGCGGC	120
Db	78	GTCGCGGCGTGCAGAGTTTACAGACAGACGCGTGTGGCGCGGGCGGGCGGTTCGCGGC	137
Qy	121	ACGTGTCTCGGACAGTGGCAGCGCGCCTGCGCCCTGGGACTTGGAGGCGCGCGCCCTGGA	180
Db	138	ACGTGTCTCGGACAGTGGCAGCGCGCCTGCGCCCTGGGACTTGGAGGCGCGCGCCCTGGA	197
Qy	181	TCCGCGGGCGCGTGGTGGCGGAGTGGGTGTGCTGCTTACCATGCGGACGCGTTCAATTGC	240
Db	198	TCCGCGGGCGCGTGGTGGCGGAGTGGGTGTGCTGCTTACCATGCGGACGCGTTCAATTGC	257
Qy	241	AGCCGCGAGAGCTCAGCTCCGACCCCGCGCGCCGCGCTTCCACGAGAGTTTCATCTG	300
Db	258	AGCCGCGAGAGCTCAGCTCCGACCCCGCGCGCCGCGCTTCCACGAGAGTTTCATCTG	317
Qy	301	CTCCTTCCCTTACCTACAGCGCCATTACAGCAAGCCTGCAAGCTTACGCGCACTGTG	360

Db	318	CTCTTCCCTGCACTGACGCGCCAAATTACACAAAGCCTGGAAGCTTGACGCCACTGTG	377
Oy	361	CAAGCACACGGGGAGAGACCAATTGTTGTGTGACTATGAAGGCTGTGGCAAGCCCTCAT	420
Db	378	CAAGGACACGGGGAGAGACCAATTGTTGTGTGACTATGAAGGCTGTGGCAAGCCCTCAT	437
Oy	421	CAGGAGCTACCATCTGAGGCCGCCACATTTCTGACTCAGACAGAGGAAAGCCGTTTGTGG	480
Db	438	CAGGAGCTACCATCTGAGGCCGCCACATTTCTGACTCAGACAGAGGAAAGCCGTTTGTGG	497
Oy	481	TGCAGCCACTGGCTGTATATCAAAAAATTCACACAAAAATCAAACTTGAAAGAAACATTTTGA	540
Db	488	TGCAGCCAAATGGCTGTATCAAAAAATTCACACAAAAATCAAACTTGAAAGAAACATTTTGA	557
Oy	541	ACGCAAAACATGAATAACAAAAACAAATATATATGCACTTTTGAAGACTGTGAAGAAC	600
Db	558	ACGCAAAACATGAATAACAAAAACAAATATATATGCACTTTTGAAGACTGTGAAGAAC	617
Oy	601	CTTTAAGAAGATAGAGAGCTGAAAAATCCATCAGTGCAGCATACCAATGAAGCCCTCAT	660
Db	618	CTTTAAGAAGATAGAGAGCTGAAAAATCCATCAGTGCAGCATACCAATGAAGCCCTCAT	677
Oy	661	CAAGTGTAACCGAGAGAGATGTGGAAACACTTTGGATCACCCAGCAAGCTGAAGACGA	720
Db	678	CAAGTGTAACCGAGAGAGATGTGGAAACACTTTGGATCACCCAGCAAGCTGAAGACGA	737
Oy	721	TGCCAAGGCCACGAGAGGCTATGTATCTCAAAAAAGATGTTCTTTGTGGCAAAAAACATG	780
Db	738	TGCCAAGGCCACGAGAGGCTATGTATCTCAAAAAAGATGTTCTTTGTGGCAAAAAACATG	797
Oy	781	GACGGAACCTTGGAACATGTGAGAGAAACCCATTAAGAGGAAATCTATGGAAGATAG	840
Db	798	GACGGAACCTTGGAACATGTGAGAGAAACCCATTAAGAGGAAATCTATGGAAGATAG	857
Oy	841	CCGGA AAAACATTTAAAGCGCAAGATTACCTTAAGCAACATGAAACATCAATGCCCCAGA	900
Db	858	CCGGA AAAACATTTAAAGCGCAAGATTACCTTAAGCAACATGAAACATCAATGCCCCAGA	917
Oy	901	AAGGATGTATGTGCTGCTGCACAGAGAGGCTGTGGAAGACCTATATCTACTGTGTTTAA	960
Db	918	AAGGATGTATGTGCTGCTGCACAGAGAGGCTGTGGAAGAACCTATATCAACACTGTGTTAA	977
Oy	961	TCTCCAAAAGCCATATCGCTCTCCCTCCATGAGGAAAGCGCCGCTTTGTGTGAACATGC	1020
Db	978	TCTCCAAAAGCCATATCTCTCTCTCCATGAGGAAAGCGCCGCTTTGTGTGAACATGC	1037
Oy	1021	TGGCTGTGGCAAAAACATTTGCAATGAACAAGCTCTACAGGCATGCTTTGTACATGA	1080
Db	1038	TGGCTGTGGCAAAAACATTTGCAATGAACAAGCTCTACAGGCATGCTTTGTACATGA	1097
Oy	1081	TCCTGCACAAAGAAATGAAGCTCAAAAGTCAAAAAATTCGTGTA AAAACGGAATTTGGC	1144
Db	1098	TCCTGCACAAAGAAATGAAGCTCAAAAGTCAAAAAATTCGTGTA AAAACGGAATTTGGC	1157
Oy	1141	CTCTCATCTCAGTGGATATATCCCTCCCAAAAGGAAAGGCAAGGCTTATCTTTGTG	1200
Db	1158	CTCTCATCTCAGTGGATATATCCCTCCCAAAAGGAAAGGCAAGGCTTATCTTTGTG	1217
Oy	1201	TCAAAACGAGAGTCCACCCAACTGTGTGAAGACAAGATGCTCTCGACAGTTGCACTACT	1266
Db	1218	TCAAAACGAGAGTCCACCCAACTGTGTGAAGACAAGATGCTCTCGACAGTTGCACTACT	1277
Oy	1261	TACCTTGGCTTAA 1273	
Db	1278	TACCTTGGCTTAA 1290	

```

RESULT 12
US-60-324-185-16829
; Sequence 16829, Application US/60324185
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti

```



```
|||||
Db 20 ATGGCGAGCGCGCGCGAGCGCGGGCGGTGCTGCTGACCGCGCGCTCCCGAGAGT 79
Qy 61 GTGGCGGCGTGGCGGGAAGGTTCAGCAGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCG 120
Db 80 GTGGCGGCGCGCGCGGAAGGTTCAGCAGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 139
Qy 121 ACCTGTCTCGGCAAGT -GGCAGCGCGCGCTGGCCCT -GGCGTTGGAGCGCGCGCGCGCGCT 178
Db 140 ACCTGTCTCGGCAAGTGGGAGCGCGCGCTGGCCCTGGCGCGTGGAGCGCGCGCGCGCGCT 199
Qy 179 GATCGCGCGCGCGCGTGGTGGCGCGAGTGGGTGTCTGCTTGAACATCGCGCAGCGCTT -CAT 237
Db 200 GATCGCGCGCGCGCGTGGTGGCGCGAGTGGGTGTCTGCTTGAACATCGCGCAGCGCTTGCAT 259
Qy 238 TGCAGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCG -TTCCAGAGAGTTTCA 296
Db 260 TGCAGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 319
Qy 297 TCTGCTCTTCCCTGACTGCGAGCGCGCAATTACAGCAAAAGCGTGGAGCTTGCAGCGCGCAC 356
Db 320 TCTGCTCTTCCCTGACTGCGAGCGCGCAATTACAGCAAAAGCGTGGAGCTTGCAGCGCGCAC 379
Qy 357 TGTGCAAGCAGCAGCGGAGAGACCATTTGTTGTGACTATGAAAGGCTGTGGCAGGCGCT 416
Db 380 TGTGCAAGCAGCAGCGGAGAGACCATTTGTTGTGACTATGAAAGGCTGTGGCAGGCGCT 439
Qy 417 TCATCAGGAGCTACCATCTGAGCGCGCGCACATTCGACTCAGCAGGAGAAAGCGCGTTTG 476
Db 440 TCATCAGGAGCTACCATCTGAGCGCGCGCACATTCGACTCAGCAGGAGAAAGCGCGTTTG 499
Qy 477 TTTGTGAGCGCAGCTGCTGTGATCAAAAATTCACACAAATCAAACTGAAGAAACATT 536
Db 500 TTTGTGAGCGCAATGGCTGTGATCAAAAATTCACACAAATCAAACTGAAGAAACATT 559
Qy 537 TTGAGCGCAAAACATGAATAATCAACAAAACATATATATCAGTTTGAAGACTGTAGA 596
Db 560 TTGAGCGCAAAACATGAATAATCAACAAAACATATATATCAGTTTGAAGACTGTAGA 619
Qy 597 AGACCTTTAAGAAACATCAGACGCTGAAATTCATGAGCGAGCATACCAATGAAACCTC 656
Db 620 AGACCTTTAAGAAACATCAGACGCTGAAATTCATGAGCGAGCATACCAATGAAACCTC 679
Qy 657 TATTCAGGTATACCCAGGAAGATGTGGAAACACTTTGCATCAGCAGCAAGCTGAAAC 716
Db 680 TATTCAGGTATACCCAGGAAGATGTGGAAACACTTTGCATCAGCAGCAAGCTGAAAC 739
Qy 717 GACTGTGCAAGGCGCAGAGGGCTATGTATGTCAAAAAGCATGCTTGTGTGCAAAA 776
Db 740 GACTGTGCAAGGCGCAGAGGGCTATGTATGTCAAAAAGCATGCTTGTGTGCAAAA 799
Qy 777 CATGAGGAGCACTTCTGAAACATGTAGAGAAACCATTAAGAGGAATACTATGTGAAG 836
Db 800 CATGAGGAGCACTTCTGAAACATGTAGAGAAACCATTAAGAGGAATACTATGTGAAG 859
Qy 837 TATGCCGGAAGAACTTTAAAGCAAGATTACCTTAAGCAACACATGAAGAACTCATGCC 896
Db 860 TATGCCGGAAGAACTTTAAAGCAAGATTACCTTAAGCAACACATGAAGAACTCATGCC 919
Qy 897 CAGAAAGGATGTATGTGCTGTCAAGAGAGGCTGTGAGAGAACTTACTACTGTGT 956
Db 920 CAGAAAGGATGTATGTGCTGTCAAGAGAGGCTGTGAGAGAACTTACTACTGTGT 979
Qy 957 TTAATCTCCAAAGCATATCTCTCTTCATGAGGAAAGCGCGCGCTTTGTGTGTAAG 1016
Db 980 TTAATCTCCAAAGCATATCTCTCTTCATGAGGAAAGCGCGCGCTTTGTGTGTAAG 1039
Qy 1017 ATGCTGGCTGTGGCAAAACATTTGCAATGAAGCAAAAGCTCTACTAGGACATGCTTTGTA 1076
Db 1040 ATGCTGGCTGTGGCAAAACATTTGCAATGAAGCAAAAGCTCTACTAGGACATGCTTTGTA 1099
Qy 1077 ATGATCTGACAGAGAAATGAGCTCAAAAGTCAAAAATCTCTGTGAAGAAAGGAGTT 1136
|||||
```

```
Db 1100 ATGATCTGACAGAGAAATGAGCTCAAAAGTCAAAAATCTCTGTGAAGAAAGGAGTT 1159
Qy 1137 TGCGCTCTCATCTCAGTGSATATATCCCTCCAAAAGGAACAAAGGCAAGGCTTATCTT 1196
Db 1160 TGCGCTCTCATCTCAGTGSATATATCCCTCCAAAAGGAACAAAGGCAAGGCTTATCTT 1219
Qy 1197 TGTGTCAAAAGCAGAGACTCCCAACTGTGTGAGAGCAAGATGCTCTCAGCAGTTTCAG 1256
Db 1220 TGTGTCAAAAGCAGAGACTCCCAACTGTGTGAGAGCAAGATGCTCTCAGCAGTTTCAG 1279
Qy 1257 TACTTACCTTGGCTAA 1273
Db 1280 TACTTACCTTGGCTAA 1296
|||||
```

RESULT 14

```
US-09-831-426-4
; Sequence 4, Application US/09831426
; GENERAL INFORMATION:
; APPLICANT: Hoechst Marion Roussel
; APPLICANT: Borden-Pallier, F.
; APPLICANT: Rocher, C.
; TITLE OF INVENTION: Human htfIIIA gene and coded htfIIIA protein
; FILE REFERENCE: 146.1364
; CURRENT APPLICATION NUMBER: US/09/831,426
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 4
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
US-09-831-426-4
```

Query Match 95.2%; Score 1211.4; DB 32; Length 1213;

Best Local Similarity 99.9%; Pred. No. 2,1e-265;

Matches 1212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 61 GTGGCGGCGTGGCGGGAAGGTTCAGCAGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 1 GTGGCGGCGCGCGCGGAGGTTCAGCAGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Qy 121 ACCTGTCTCGGCAAGTGGCAGCGCGCGCTGGCCCTGGGCTTGGAGGCGCGCGCGCGCTGGA 180
Db 61 ACCTGTCTCGGCAAGTGGCAGCGCGCGCGCTGGCCCTGGGCTTGGAGGCGCGCGCGCGCTGGA 120
Qy 181 TCGCGCGCGCGTGGTGGCGGAGTGGGTGTGCTGCTTACCATGCGGAGCGCTTACATTC 240
Db 121 TCGCGCGCGCGTGGTGGCGGAGTGGGTGTGCTGCTTACCATGCGGAGCGCGCTTACATTC 180
Qy 241 AGCGGCGGAGGCTCAGCTCGACCGCGCGCGCGCGCGCGCGCTTCCAGAGGTTCAATCTG 300
Db 181 AGCGGCGGAGGCTCAGCTCGACCGCGCGCGCGCGCGCGCGCGCTTCCAGAGGTTCAATCTG 240
Qy 301 CTCCTTCCCTGACTGCAAGCCCAATTAACAGCAAAAGCTTGAAGCTTGAACGCACTGTG 360
Db 241 CTCCTTCCCTGACTGCAAGCCCAATTAACAGCAAAAGCTTGAAGCTTGAAGGCACTGTG 300
Qy 361 CAAGCAGCGGCGGAGAGACCAATTTGTTGTGACTATGAAGGCTGTGGCAAGCGCTTCA 420
Db 301 CAAGCAGCGGCGGAGAGACCAATTTGTTGTGACTATGAAGGCTGTGGCAAGCGCTTCA 360
Qy 421 CAGGAGTACCATCTGAGCGCGCAATTCGACATCAGCAGGAGAAAGCGTTGTGTTG 480
Db 361 CAGGAGTACCATCTGAGCGCGCAATTCGACATCAGCAGGAGAAAGCGTTGTGTTG 420
Qy 481 TGCAGCAGCTGGCTGTGATCAAAAATTCACACAAATTCAACTTGAAGAAACATTTTGA 540
Db 421 TGCAGCAGCTGGCTGTGATCAAAAATTCACACAAATTCAACTTGAAGAAACATTTTGA 480
Qy 541 ACGCAAAACATGAATAATCAACAAAACATATATATGACAGTTTGAAGACTGTGAAGAG 600
Db 481 ACGCAAAACATGAATAATCAACAAAACATATATATGACAGTTTGAAGACTGTGAAGAG 540
|||||
```



```

|||||
Db 1081 CTCTCATCTCAGTGGATATATCCCTCCAAAGGAAACAAGGCAAGGCTTATCTTTGTG 1140
Qy 1201 TCAAAACGGAGACTCACCCAACTGTGTGGAGACAAAGATGCTCTCGACAGTTGCAGTACT 1260
Db 1141 TCAAAACGGAGACTCACCCAACTGTGTGGAGACAAAGATGCTCTCGACAGTTGCAGTACT 1200
Qy 1261 TACCCTTGGCTAA 1273
|||||
Db 1201 TACCCTTGGCTAA 1213

```

Search completed: February 10, 2003, 14:48:16
Job time : 3133.12 secs